

# ARNOLD DIFFUSION IN NEARLY INTEGRABLE HAMILTONIAN SYSTEMS OF ARBITRARY DEGREES OF FREEDOM

CHONG-QING CHENG AND JINXIN XUE

ABSTRACT. In this paper Arnold diffusion is proved to be a generic phenomenon in nearly integrable convex Hamiltonian systems with arbitrarily many degrees of freedom:

$$H(x, y) = h(y) + \varepsilon P(x, y), \quad x \in \mathbb{T}^n, \quad y \in \mathbb{R}^n, \quad n \geq 3.$$

Under typical perturbation  $\varepsilon P$ , the system admits “connecting” orbit that passes through any finitely many prescribed small balls in the same energy level  $H^{-1}(E)$  provided  $E > \min h$ .

## CONTENTS

1. Introduction	2
1.1. Statement of the main result	4
1.2. Convention of notations and terminology	8
2. The normal form	10
2.1. Number theoretic properties of the frequency line	10
2.2. Resonant submanifolds and their neighborhoods	12
2.3. Homogenization	13
2.4. The normal form around single resonance	15
2.5. The normal form around double resonance	16
3. The choice of diffusion path: from single, double to complete resonances	17
3.1. Normally hyperbolic invariant manifold for Hamiltonian system	17
3.2. Normally hyperbolic invariant cylinder (NHIC) in the regime of single resonances	18
3.3. Strong double resonances and shear transformation	20
3.4. NHIC in Hamiltonian systems of two degrees of freedom	22
3.5. NHIC in the intermediate region between single and strong double resonance	24
3.6. Center straightening and reduction of Hamiltonian systems with two degrees of freedom	29
3.7. Reduction of order and undoing the shear near strong double resonance in the full system	34
3.8. Frequency tracking and refinement	36
3.9. Normal form after one step of order reduction around strong double resonance	38
3.10. Reduction of order around triple resonance	39
3.11. Further genericity conditions	43
4. A hierarchy of canonical transformations	45

4.1.	Induction at the single and weak double resonances	45
4.2.	Induction around the strong double resonances	48
4.3.	Reduction around complete resonance	52
4.4.	Admissible frequency path	53
5.	Mañé sets along $s$ -resonance and around $m$ -strong resonance	56
5.1.	Mañé set along $s$ -resonance	56
5.2.	The flat $\mathbb{F}_0$ of the $\alpha$ -function	58
5.3.	The channel connected to $\mathbb{F}_0$	61
6.	How to cross $m$ -strong resonances	62
6.1.	The difficulty and the strategy	62
6.2.	Cohomology equivalence around $m$ -strong resonance	63
6.3.	Overlapping property	66
6.4.	Construction of the ladder	67
7.	Variational construction of global diffusion orbits	72
7.1.	Local connecting orbit of type- $h$ with incomplete intersections	72
7.2.	Local connecting orbits of type- $c$	78
7.3.	Existence of generalized transition chain	80
7.4.	Variational construction of globally connecting orbits	81
8.	Genericity and the proof of the main theorem	82
8.1.	Generic conditions: the multi-filtered cusp-residual condition (MFCR)	83
8.2.	Proof of the main theorem	86
9.	Parametrization of weak KAM solutions and its modulus of continuity	88
9.1.	Topology of level set of weak KAM solutions	88
9.2.	Modulus of continuity of weak KAM solutions	91
9.3.	Extension of weak KAM solutions using the normal hyperbolicity	99
	Appendix A. A brief introduction to Mather theory and weak KAM	100
A.1.	Minimizing measure and $\alpha, \beta$ function	100
A.2.	(Semi)-static curves, the Aubry set and the Mañé set	101
A.3.	Elementary Weak KAM	103
	References	108

## 1. INTRODUCTION

In this paper, we consider nearly integrable Hamiltonian systems of the form

$$(1.1) \quad H(x, y) = h(y) + \varepsilon P(x, y), \quad (x, y) \in T^*\mathbb{T}^n, \quad n \geq 3.$$

where  $h$  is strictly convex, namely, the Hessian matrix  $\frac{\partial^2 h}{\partial y^2}$  is positive definite. It is also assumed that both  $h$  and  $P$  are  $C^r$ -function with  $r > 2n$  and  $\min h = 0$ .

The problem of studying the (in)stability of the above system  $H$  was considered to be the fundamental problem of Hamiltonian dynamics by Poincaré. According to the celebrated KAM theorem, there exists a large measure Cantor set of Lagrangian tori on which the dynamics is conjugate to irrational rotations and the oscillation of the slow variable (or called action variable)  $y$  is at most  $O(\sqrt{\varepsilon})$ . The KAM theorem also excludes the possibility of large oscillation of  $y$  in the case of  $n = 2$  since each energy level, which is three dimensional, is laminated by two dimensional KAM tori and each orbit either stays on a KAM torus or is confined between two tori.

For  $n \geq 3$ , there does not exist topological obstruction for the slow variables  $y$  to have  $O(1)$  oscillation. Arnold was the first one who had realized such instability [A63] and constructed the first example in [A64] half a century ago

$$(1.2) \quad H(I, \theta, y, x, t) = \frac{I^2}{2} + \frac{y^2}{2} + \varepsilon(\cos x + 1)(1 + \mu(\cos \theta + \sin t)),$$

where there are orbits along which the action variable  $I$  has as large oscillation as we wish. Although the perturbation is far from being typical, Arnold still proposed

**Conjecture** ([A66]): *The “general case” for a Hamiltonian system (1.1) with  $n \geq 3$  is represented by the situation that for an arbitrary pair of neighborhood of tori  $y = y'$ ,  $y = y''$ , in one component of the level set  $h(y) = h(y')$  there exists, for sufficiently small  $\varepsilon$ , an orbit intersecting both neighborhoods.*

The problem is named Arnold diffusion. In recent years, it has become clear that Arnold diffusion is a typical phenomenon in so called *a priori* unstable systems, which are small perturbation of compound pendulum-single rotator system. There are many works studying this problem based on two streams of methods: the variational method (c.f. [Be2, CY1, CY2, LC]) and the geometric method (c.f. [DLS06, DLS13, Tr]). With the variational method, the genericity of perturbations was established in [CY1, CY2], which relies on the existence of a normally hyperbolic invariant cylinder (NHIC) of dimension two given by the *a priori* unstable condition. For the geometric method, we highlight the central object, the scattering map, whose symplectic properties are studied in details in [DLS08].

Nearly integrable Hamiltonian like (1.1) is also called *a priori* stable system, where a frequency  $\omega(y) = \frac{\partial h}{\partial y}$  is said to admit a resonance relation, if there exists an integer vector  $\mathbf{k} \in \mathbb{Z}^n \setminus \{0\}$  such that  $\langle \mathbf{k}, \omega(y) \rangle = 0$  at some point  $y$ . The number of linearly independent resonance relations is called the multiplicity of the resonance. A bit away from strong multiple resonance in such systems, some pieces of NHIC still exist and the method for *a priori* unstable system can still be applied [Be3, BKZ]. However, these pieces of NHIC are separated by small neighborhood of multiple resonances. Without studying how to pass through neighborhood of strong multiple resonances, it would be impossible to construct orbits which can drift for large scale.

The problem of Arnold diffusion has a long history, and the mathematical issues that arise are subtle and technically involved. Many key ideas can be traced back to the early published and unpublished work of Mather (see the announcement [M04] in 2004 and published errata [M12] in 2012, as well as the paper [M11]). Even in the case  $n = 3$ , there are (as of the date of this paper) no published proofs of diffusion. There are, however, several papers on the arxiv claiming the result for  $n = 3$ : a paper of Cheng was first posted in Jul. 2012 (updated in Mar. 2013, later decomposed into [C15a, C15b, CZ1] in 2015 and [CZ1] is accepted); between the first and second version of Cheng’s paper, Kaloshin and K. Zhang posted a paper [KZ1] on arxiv in Dec. 2012

(updated in Jan 2013 and a published announcement [KZ4]); Marco and Sabbagh posted partial solutions [MaSa] on arxiv in Jan. 2014 (see also [Mc] a conference paper in 2015).

In April of 2014, Kaloshin and Zhang posted on their webpages a 36-page announcement and key elements of a proof of diffusion [KZ2] in the case  $n = 3.5$ . Subsequently, they posted on the arxiv in October 2014 a longer paper [KZ3] containing details of the announced proof as well as partial results and techniques applicable to  $n \geq 4$ . This paper bears some important commonalities with part of our paper. We discuss these, as well as the crucial new ideas in this paper including issues of crossing complete resonance, in the next section. The present paper is the first posted complete proof of Arnold diffusion for general  $n \geq 4$ .

**1.1. Statement of the main result.** By adding a constant to  $H$  and introducing a translation  $y \rightarrow y + y_0$ , one can assume  $\min h(y) = h(0) = 0$ . For  $E > 0$ , let  $H^{-1}(E) = \{(x, y) : H(x, y) = E\}$  denote the energy level set,  $B \subset \mathbb{R}^n$  denote a ball in  $\mathbb{R}^n$  such that  $\bigcup_{E' \leq E+1} h^{-1}(E') \subset B$ . Let  $\mathfrak{S}_a, \mathfrak{B}_a \subset C^r(\mathbb{T}^n \times B)$  denote a sphere and a ball with radius  $a > 0$  respectively:  $F \in \mathfrak{S}_a$  if and only  $\|F\|_{C^r} = a$  and  $F \in \mathfrak{B}_a$  if and only  $\|F\|_{C^r} \leq a$ . They inherit the topology from  $C^r(\mathbb{T}^n \times B)$ .

For perturbation  $P$  independent of  $y$  (for instance, in classical mechanical systems), we use the same notation  $\mathfrak{S}_a, \mathfrak{B}_a \subset C^r(\mathbb{T}^n)$  to denote a sphere and a ball with radius  $a > 0$ .

Let  $\mathfrak{R}_a$  be a set residual in  $\mathfrak{S}_a$ , each  $P \in \mathfrak{R}_a$  is associated with a set  $R_P$  residual in the interval  $[0, a_P]$  with  $a_P \leq a$ . A set  $\mathfrak{C}_a$  is said cusp-residual in  $\mathfrak{B}_a$  if

$$\mathfrak{C}_a = \{\lambda P : P \in \mathfrak{R}_a, \lambda \in R_P\}.$$

Let  $\Phi_H^t$  denote the Hamiltonian flow determined by  $H$ . Given an initial value  $(x, y)$ ,  $\Phi_H^t(x, y)$  generates an orbit of the Hamiltonian flow  $(x(t), y(t))$ . An orbit  $(x(t), y(t))$  is said to visit  $B_\delta(y_0) \subset \mathbb{R}^n$  if there exists  $t \in \mathbb{R}$  such that  $y(t) \in B_\delta(y_0)$  a ball centered at  $y_0$  with radius  $\delta$ .

**Theorem 1.1.** *Given any small  $\delta > 0$ , there exists  $\varepsilon_0$ , such that given finitely many small balls  $B_\delta(y_i) \subset \mathbb{R}^n$ , where  $y_i \in h^{-1}(E)$  with  $E > 0$ , there exists a cusp-residual set  $\mathfrak{C}_{\varepsilon_0} \subset C^r(\mathbb{T}^n \times B)$  with  $r > 2n$  such that for each  $\varepsilon P \in \mathfrak{C}_{\varepsilon_0}$ , the Hamiltonian flow  $\Phi_H^t$  admits orbits which visit the balls  $B_\delta(y_i)$  in any prescribed order. Moreover, the theorem still holds if we replace the function space  $C^r(\mathbb{T}^n \times B)$  by  $C^r(\mathbb{T}^n)$ .*

This is the main result of this paper. Our statement is stronger than the original conjecture of Arnold stated above. We establish the existence of  $\delta$ -dense orbits in the phase space for any given  $\delta$ . We have the following remarks.

(1) **Potential or metric perturbations.** Our perturbation is generic not only in usual sense, but also in the sense of Mañé [Me], i.e. it is a typical phenomenon when the system is perturbed by functions depending only on  $x$ , called potential perturbation. It is interesting to ask the following *Riemannian metric version of Arnold diffusion* (see [KL] for an example).

**Problem:** Suppose  $g = (g_{ij}^0(x) + \varepsilon f_{ij}(x))dx^i dx^j$ ,  $x \in \mathbb{T}^n$  and  $i, j = 1, \dots, n$ , is a  $C^r$ ,  $r > 2n$  Riemannian metric on  $\mathbb{T}^n$ ,  $n \geq 3$  where  $g_{ij}^0(x)dx^i dx^j$  is flat. Prove that for “generic”  $f = (f_{ij})$  in the unit sphere of  $(C^r(\mathbb{T}^n))^{n \times n}$ , any two small balls in the unit tangent bundle  $T^1\mathbb{T}^n$  are connected by an orbit  $(\gamma, \dot{\gamma})$ , where  $\gamma$  is a geodesic, for sufficiently small  $\varepsilon$ .

(2) **The cusp-residual condition.** Roughly speaking, cusp-residual means that for generic  $P \in \mathfrak{S}_1$ , there is  $\varepsilon_P$  depending on  $P$  such that (1.1) admits diffusing orbit for generic  $\varepsilon < \varepsilon_P$ , but for given  $\varepsilon$ , there is an open set on  $\mathfrak{S}_\varepsilon$  for  $\varepsilon P$  therein the existence of diffusing orbit is not known. Our cusp-residual set has an extra structure called multiple filtration described in Section 8, so we call it a multi-filtered cusp-residual (MFCR) set.

(3) **The smoothness of the perturbations.** The smoothness requirement  $r > 2n$  matches the least KAM smoothness threshold. Namely, it was shown in [C11] that KAM theory does not apply if the perturbation is small only in  $C^{2n-\delta}$  but not  $C^{2n}$ -topology (see Remark 4.1). Our theorem holds also for  $r = \infty$ , but the analytic perturbation case of Arnold diffusion remains open (see [GT2] for a result on *a priori* chaotic systems). On the other hand, for Hamiltonians with low regularity, connecting lemma techniques can be applied to get topologically mixing for “most Hamiltonian” and “most energy levels” [BFRV]. See also [ABC, BoCr] for other related results on  $C^1$  generic symplectomorphisms. Robust transitivity is shown in [NP] to be true for an open set of Hamiltonians in  $C^r$ ,  $r > 2$  by introducing a symplectic blender. For non Hamiltonian but volume preserving  $C^1$  generic diffeomorphisms, it is shown in [ACW] that there is a dichotomy, i.e. either zero metric entropy or the volume being ergodic.

This theorem is proved by variational method. The diffusion orbits are constructed to **move along many pieces of NHIC** and to **cross multiple-strong resonance point** (join two pieces of NHIC).

**1.1.1. Outline of our proof.** The construction of diffusing orbits involves understanding the pictures in four different spaces: the frequency space, the phase space, the parameter space (or the cohomology space  $H^1(\mathbb{T}^n, \mathbb{R})$ ) and the function space, of which we have a clear understanding summarized as follows.

(1) The frequency space picture.

- It is a folklore theorem that Fourier modes in the  $\text{span}_{\mathbb{Z}}\{\mathbf{k}\}$  appears in the KAM normal form if  $\mathbf{k}$  is a resonance relation to the frequency  $\omega(y)$  (Section 2). In other words, “resonance produces pendulum”.
- We show that for any two balls in the frequency space, there is a connected curve joining them with a hierarchy structure (Section 4.4). Namely, along the curve each frequency has at least  $(n-2)$  linearly independent resonant integer vectors  $\mathbf{k}^1, \mathbf{k}^2, \dots, \mathbf{k}^{n-2} \in \mathbb{Z}^n$  forming a hierarchy  $|\mathbf{k}^i| \ll |\mathbf{k}^{i+1}|$ ,  $i = 1, \dots, n-3$  except for finitely many points, there are  $(n-1)$  linearly independent resonant integer vectors  $\mathbf{k}^1, \mathbf{k}^2, \dots, \mathbf{k}^{n-1} \in \mathbb{Z}^n$  forming a hierarchy  $|\mathbf{k}^i| \ll |\mathbf{k}^{i+1}|$ ,  $i \neq j$ , for some  $j = 1, 2, \dots, n-2$  and  $|\mathbf{k}^j|$  is comparable to  $|\mathbf{k}^{j+1}|$ .

(2) The phase space picture.

- The orbits always shadow certain Aubry-Mather set for twist maps or subsystems of two degrees of freedom.
- The (un)stable “manifolds” of these Aubry-Mather set do not always need to intersect *transversally* in order to implement Arnold’s mechanism. In some circumstance, it is enough to have incomplete intersections (see Theorem 7.1), namely they do not split along some but not all directions.

(3) The parameter space (the cohomology space  $H^1(\mathbb{T}^n, \mathbb{R})$ ) picture.

- Based on the complete understanding of the flat of the  $\alpha$  function for systems of two degrees of freedom in [C12, C15b]. We get a continuous path in  $H^1(\mathbb{T}^n, \mathbb{R})$  along which the  $c$ -equivalence is always satisfied (Section 6.2).
- The difficulty of the high dimensionality manifests itself as the channel misalignment problem (see Fig 2 and compare Fig 1). We address it by introducing a ladder construction (Section 6.4). The corresponding picture in the frequency path has a detour of order  $\sqrt{\varepsilon}$  close to strong double resonance (see Fig 5).

(4) The function space picture.

- We get a MFCR set using the parametric transversality (PT) results established in [CZ1, CZ2, CY1, CY2], perturbations in which admit diffusing orbit (Section 8).
- We verify the assumptions for the PT results by showing that the weak KAM solutions for systems with two degrees of freedom restricted to the energy level can be parametrized to have  $1/3$  Hölder regularity with respect to the parameter. This is used in the ladder construction.

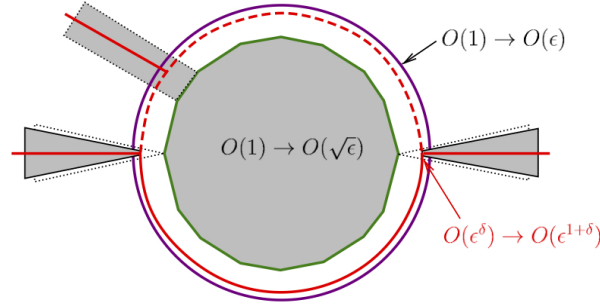


FIGURE 1. The  $n = 3$  case, (red) curves of cohomology equivalence

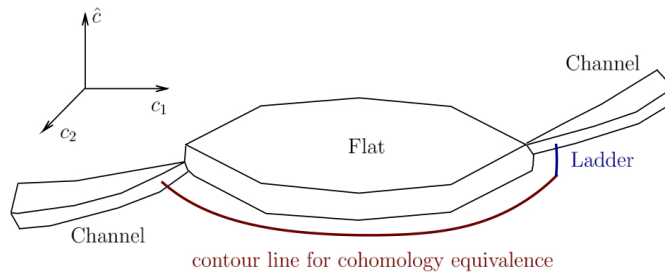


FIGURE 2. The  $n > 3$  case, the pizza and the ladder climbing

Most of the above outlined work is devoted to understanding the dynamics near complete resonance where the frequency  $\omega(y)$  admits  $n - 1$  resonance relations two of which are strong. The main difficulty comes from the nonperturbative nature of the dynamics near strong resonance which is intrinsic to the *a priori* stable systems. Our work on crossing the complete resonance may have broader interests beyond Arnold diffusion. Away from the strong resonances along the frequency line, NHICs can be found so that the problem is reduced to the well understood *a priori* unstable system.

We next briefly explain the construction of NHICs approaching the multiple-strong resonant points. A general idea is that sufficiently weak resonance can be averaged and

treated as small perturbation to those stronger resonance terms. In the earlier papers [KZ2, KZ3] this idea is used, where the “dominant structure” is introduced to find a path of NHIC. Independently, we work out this idea in this paper in the following way. Inspired by Arnold’s remark in [A66] “it is necessary to examine the transition from single to double resonance” we choose the frequency line  $\omega_a = \lambda_a(a, \frac{P}{Q}, \frac{p}{q}, \hat{\omega}_{n-3}^*)^t$  carefully, where  $\hat{\omega}_{n-3}^* \in \mathbb{R}^{n-3}$  is of Diophantine. When the parameter  $a$  varies in an interval, we get a path in  $h^{-1}(E)$  by choosing suitable  $\lambda_a$ .

- Two types of KAM iteration are introduced so that we reduce the system (1.1) into a locally defined system with one less degree of freedom for generic Hamiltonian. It is guaranteed by a result [DLS08] saying that a Hamiltonian system restricted on the NHIM is still Hamiltonian. Restricted on NHIM, the frequency vector turns out to be  $\omega'_a = \lambda_a(a, 0, \frac{d_0}{qQ}, \hat{\omega}_{n-3}^*)^t$  after transformed by unimodular matrix. The zero entry in  $\omega'_a$  corresponds to the normal dimension, which is eliminated by the reduction. Next, we approximate  $\hat{\omega}_{n-3}^*$  by a vector  $(\frac{\bar{p}}{q}, \hat{\omega}_{n-4}^*)$ , where  $\hat{\omega}_{n-4}^*$  is of Diophantine such that the new line segment  $\bar{\omega}_a = \lambda_a(a, \frac{d_0}{qQ}, \frac{\bar{p}}{q}, \hat{\omega}_{n-4}^*)^t$  has the same structure as the  $\omega_a$  before.
- Around strong double resonant point, extra action-angle coordinates are introduced (center straightening) so that the reduced system is nearly integrable again. Such coordinate transformation is singular only at the double resonant point. It allows us to repeat the procedure of reduction in place arbitrarily close to the multiple-strong resonance points.
- In our reduction of order, we obtain explicitly Hamiltonian systems of one less degrees of freedom and during each reduction of order we have at most two resonances to consider.

Such procedure is repeated for  $n - 3$  steps, certain NHICs are found to reach  $\mu\sqrt{\varepsilon}$  neighborhoods of multiple-strong resonant points with  $\mu \ll 1$ . Once NHICs are obtained, we can construct diffusion orbits along these NHICs by the method developed in [CY1, CY2, C12].

The choice of path for NHIC in [KZ2, KZ3] is top-down. They first show that there exists “a connected  $\rho$ -dense admissible diffusion path”, by considering all resonant integer vectors satisfying certain assumptions, then choose the path implicitly from this connected set. Our approach is bottom-up. We first choose our diffusing path (frequency line segment) explicitly then determine all its resonance integer vectors explicitly. For any given two balls in the frequency space, we find a path connecting two rational points in the balls. To see the picture clearly, let us map the frequency path to  $\mathbb{R}^n$  by setting  $\lambda_a = 1$ . Then the two rational points, as two opposite vertices, determines a cuboid. Our path connects the two opposite vertices moving along the edges of the cuboid and has the hierarchy structure described before.

**1.1.2. Structure of the paper.** The paper is organized as follows. In Section 2, to get two normal forms corresponding to single or double resonances we design two types of KAM iteration by applying a proceeding homogenization procedure, which are used to construct two type of reductions of order in Section 3 correspondingly. One is for single and weak double resonance, another one is for strong double resonance so that we get a Hamiltonian system with one less degrees of freedom. Such construction is based on normally hyperbolic invariant manifold theorems and its Hamiltonian version following [DLS08]. Special efforts are made for strong double resonance, includes

locating of NHICs around the strong double resonance following [C12, CZ1], introducing new action-angle coordinates and performing shear transformation. In Section 4, we show how to perform the procedure of reduction of order of Section 3 inductively to get NHICs of two dimensions. We also explain how to connect two  $\rho$  balls in the frequency space by frequency lines of the standard form  $\omega_a$ . In Section 5, we give a description of the Aubry set along the frequency line segment and around multiple-strong resonant points. In Section 6.2, we show how to pass through multiple-strong resonance, it involves “ladder climbing” to overcome the difficulty created by the triple or more multiple-strong resonance, building up the cohomology equivalence which is used to turn around the multiple-strong resonance. In Section 7, we show how to construct diffusing orbits variationally. In Section 8, we recapitulate all the genericity assumptions that we made and give a proof of the main theorem. In Section 9, we establish a Hölder regularity result for the parametrized weak KAM solutions, which is needed to complete the genericity argument of ladder construction. Finally, we give the preliminaries of the Mather theory and weak KAM in the Appendix A.

**1.2. Convention of notations and terminology.** In this section, we fix some standing conventions of notations for the rest of the paper. Please refer to Appendix A for a brief introduction of the Mather theory where more notations are given.

**1.2.1. Legendre transform.** The Lagrangian  $L$  is related to the Hamiltonian  $H$  through Legendre transformation  $H(x, y) = \max_{\dot{x}} \langle \dot{x}, y \rangle - L(x, \dot{x})$ . They produce the transformation  $\mathcal{L}_L: TM \rightarrow T^*M$ ,  $M = \mathbb{T}^n$

$$(x, \dot{x}) \rightarrow (x, y) = \mathcal{L}_L(x, \dot{x}) = \left( x, \frac{\partial L}{\partial \dot{x}}(x, \dot{x}) \right).$$

Its inverse  $\mathcal{L}_L^{-1} = \mathcal{L}_H: T^*M \rightarrow TM$  is determined by the Hamiltonian  $H$ ,

$$(x, y) \rightarrow (x, \dot{x}) = \mathcal{L}_H(x, y) = \left( x, \frac{\partial H}{\partial y}(x, y) \right)$$

Denote  $\Phi_H^t: T^*M \rightarrow T^*M$  the Hamiltonian flow produced by the function  $H$  and  $\phi_L^t: TM \rightarrow TM$  the Lagrange flow produced by the function  $L$ . If  $L$  is related to  $H$  via Legendre transformation, then

$$\mathcal{L}_L \phi_L^t = \Phi_H^t, \quad \mathcal{L}_H \Phi_H^t = \phi_L^t.$$

Let  $\alpha_L$  denote the  $\alpha$ -function for the Lagrangian  $L$ . If the Lagrangian  $L$  is obtained from the Hamiltonian  $H$  through Legendre transformation, we use  $\alpha_H$  to denote the  $\alpha$ -function, i.e.  $\alpha_H = \alpha_L$ . Similarly,  $\beta_L = \beta_H$  denotes the  $\beta$ -function for the Lagrangian  $L$  and for the Hamiltonian  $H$  respectively.

Each Tonelli Lagrangian  $L$  induces a relation between the first cohomology group and the first homology group  $\mathcal{L}_{\beta_L}: H_1(M, \mathbb{R}) \leftrightarrow H^1(M, \mathbb{R})$ , called Fenchel-Legendre transformation:

$$c \in \mathcal{L}_{\beta_L}(\rho) \iff \alpha_L(c) + \beta_L(\rho) = \langle c, \rho \rangle.$$

If a Hamiltonian  $H$  is related to a Lagrangian  $L$  via Legendre transformation, we can also use the notation  $\mathcal{L}_{\beta_H}$ , i.e.  $\mathcal{L}_{\beta_H} = \mathcal{L}_{\beta_L}$ .



1.2.2. *c-minimal curve and c-minimal orbit.* A curve  $\gamma: \mathbb{R} \rightarrow M$  is called *c-minimal* if for any curve  $\xi: \mathbb{R} \rightarrow M$  and for any  $t_0, t_1, t'_1 \in \mathbb{R}$  with  $t'_1 = t_1 \pmod{1}$  one has

$$\int_{t_0}^{t_1} (L(\gamma(t), \dot{\gamma}(t), t) - \langle c, \dot{\gamma}(t) \rangle + \alpha_L) dt \leq \int_{t_0}^{t_1} (L(\xi(t), \dot{\xi}(t), t) - \langle c, \dot{\xi}(t) \rangle + \alpha_L) dt,$$

where the Lagrangian  $L$  is assumed time-1-periodic:  $L(\cdot, t) = L(\cdot, t+1)$ . If a curve  $\gamma$  is *c-minimal*, then  $(\gamma, \dot{\gamma})$  ( $\mathcal{L}_L(\gamma, \dot{\gamma})$ ) is called *c-minimal orbit*.

1.2.3.  *$\lambda g$ -minimal periodic curve and  $\lambda g$ -minimal periodic orbit.* Consider a Tonelli Lagrangian  $L(x, \dot{x})$  independent of time defined on  $TM$ . A periodic curve  $\gamma: [0, \lambda^{-1}] \rightarrow M$  is associated with a class  $[\gamma] \in H_1(M, \mathbb{Z})$ . It is called  *$\lambda g$ -minimal periodic curve* if one has

$$\int_0^{\frac{1}{\lambda}} L(\gamma(t), \dot{\gamma}(t)) dt = \inf_{[\xi]=g} \int_0^{\frac{1}{\lambda}} L(\xi(t), \dot{\xi}(t)) dt.$$

In this case  $(\gamma, \dot{\gamma})$  ( $\mathcal{L}_L(\gamma, \dot{\gamma})$ ) is called  *$\lambda g$ -minimal orbit*.

**Notations** We list some conventions and notations.

- (*The vector norm*) Our convention of using  $|\cdot|$  as follows.
  - \* It is the usual absolute value when applied to real numbers.
  - \* It is the  $\ell_1$  norm when applied to integer vectors  $\mathbf{k} \in \mathbb{Z}^n$  which are row vectors.
  - \* It is the  $\ell_\infty$  norm when applied to a frequency  $\omega \in \mathbb{R}^n$  which is column vector.

So we can write estimate  $|\langle \mathbf{k}, \omega \rangle| \leq |\mathbf{k}| \cdot |\omega|$ .

- Let  $\sigma(i, j)$  be the permutation matrix, i.e. its  $(i, j)$ -entry,  $(j, i)$ -entry and  $(\ell, \ell)$ -entry ( $\ell \neq i, j$ ) are equal to 1, all other entries are equal to zero. When it applies to an  $n \times n$  matrix, it is also thought as  $n \times n$  matrix by default.
- (*The projection operator*) Let  $\pi_{-i}$  be an operator, when it applies to a vector, it eliminate the  $i$ -th entry, for instance,  $\pi_{-2}(v_1, v_2, v_3, \dots, v_n) = (v_1, v_3, \dots, v_n)$ . When it applies to a matrix, it eliminate the  $i$ -th row and the  $i$ -th column. The operator  $\pi_{+i}$  is introduced by adding 0 to a vector as the  $i$ -th entry, for instance,  $\pi_{+2}(v_1, v_3, \dots, v_n) = (v_1, 0, v_3, \dots, v_n)$ . For a matrix  $M \in \mathbb{R}^{j^2}$ ,  $j \geq 1$ , we define  $\pi_{+i}(M) \in \mathbb{R}^{(j+1)^2}$  as the enlarged matrix obtained by putting zeros as the new  $i$ -th row and column except for the  $(i, i)$  entry, we put 1.
- (*the hat notation*) We fix the meaning of the *hat* notation throughout the paper. For a vector  $v \in \mathbb{R}^n$ , we use  $\hat{v}_{n-i}$  to denote the vector  $(v_{i+1}, v_{i+2}, \dots, v_n)$  in  $\mathbb{R}^{n-i}$  for  $1 < i < n$ .
- (*the tilde notation*) Dual to the *hat* notation, we introduce the *tilde* notation. For a vector  $v \in \mathbb{R}^n$ , we use  $\tilde{v}_i$  to denote the vector  $(v_1, \dots, v_i) \in \mathbb{R}^i$ ,  $3 < i < n$ . We omit the subscript  $i$  if  $i = 2$ .
- (*the ' and '' notations*) We use  $'$  to indicate the single resonance and  $''$  for the double resonance (see below). We never use them to take derivatives in this paper.
- (*a number as super- or sub-scripts*) A vector may carry numeric super- or sub-scripts. The superscript counts the step of reduction of order and the subscript indicates the vector entries. Such a superscript for a vector has no danger of confusion with a power. If a matrix or a scalar carries a numeric subscript, we fix its meaning as counting the step of reduction of order and its numeric super-script is the usual power.

## 2. THE NORMAL FORM

In this section, we construct the frequency vectors with special number theoretic properties and derive Hamiltonian normal forms associated to such frequency vectors.

**Definition 2.1.** We say a Diophantine vector  $v \in \mathbb{R}^{d-1}$  is in  $\text{DC}(d, \alpha, \tau)$  if we have

$$(2.1) \quad |\langle (1, v), \mathbf{k} \rangle| \geq \frac{\alpha}{|\mathbf{k}|^\tau}, \quad \mathbf{k} \in \mathbb{Z}^d \setminus \{0\}.$$

**2.1. Number theoretic properties of the frequency line.** What we are going to prove here will be used in the following KAM iterations.

**2.1.1. Single resonance.** Consider frequency vector  $\omega_a \in \mathbb{R}^n$  of the form

$$(2.2) \quad \omega_a = \lambda_a \left( a, \frac{P}{Q}, \frac{p}{q}, \hat{\omega}_{n-3}^* \right)^t, \quad P, Q, p, q \in \mathbb{Z}, \quad a \in \mathbb{R}, \quad \hat{\omega}_{n-3}^* \in \text{DC}(n-2, \alpha, \tau).$$

To determine the parameter  $\lambda_a$ , we first think  $\omega_a$  as a point in the projective space  $\mathbb{RP}^{n-1}$  so it is now determined up to a nonzero scalar multiple, which does not influence the resonance relations. Since we know that  $y$  lies on an energy level  $E$  and since the energy hyper surface  $h^{-1}(E)$  encloses a convex set containing the origin, the equation  $h(\omega^{-1}(\omega_a)) = E$ ,  $\omega(y) = \partial_y h(y)$ , determines uniquely the scalar multiple denoted by  $\lambda_a$ . For example, when  $h(y) = \frac{1}{2} \|y\|_{\ell_2}^2$ , we see easily that  $\lambda_a = \frac{\sqrt{2E}}{\left\| \left( a, \frac{P}{Q}, \frac{p}{q}, \hat{\omega}_{n-3}^* \right) \right\|_{\ell_2}}.$

Since we assume  $\hat{\omega}_{n-3}^* \in \text{DC}(n-2, \alpha, \tau)$ , we have at most two resonances as  $a$  varies in an interval. We always have a first resonance given by the integer vector

$$\mathbf{k}' = (0, Qp, -qP, \hat{0}_{n-3})/d, \quad d := \text{g.c.d.}(pQ, Pq).$$

The g.c.d. of all the components of  $\mathbf{k}'$  is 1. Then we have

$$(2.3) \quad \begin{bmatrix} 1 & 0 & 0 & \hat{0}_{n-3} \\ 0 & \frac{Qp}{d} & -\frac{qP}{d} & \hat{0}_{n-3} \\ 0 & r & s & \hat{0}_{n-3} \\ \hat{0}_{n-3} & \hat{0}_{n-3} & \hat{0}_{n-3} & \text{id}_{n-3} \end{bmatrix} \begin{bmatrix} a \\ \frac{P}{Q} \\ \frac{p}{q} \\ \hat{\omega}_{n-3}^* \end{bmatrix} = \begin{bmatrix} a \\ 0 \\ \frac{d}{qQ} \\ \hat{\omega}_{n-3}^* \end{bmatrix}.$$

where  $r, s$  are such that  $sQp + rqP = d$ . We denote the  $n \times n$  matrix by  $M'$ , which is in  $SL(n, \mathbb{Z})$ .

**2.1.2. Double resonance, away from triple or higher resonances.** In this section, we consider that the vector (2.2) at double resonance. We fix some large number  $K$ . As  $a$  varies in an interval, we may encounter *double resonant* points

$$\{\omega_a \mid \langle \mathbf{k}, \omega_a \rangle = 0, \text{ for some } \mathbf{k} \in \mathbb{Z}^n \setminus \text{span}_{\mathbb{Z}} \{\mathbf{k}'\}, \text{ and } |\mathbf{k}| \leq K\}.$$

There are finitely many such double resonant points, whose number depends only on  $K$ .

The next lemma shows that for fixed  $K$ , points along the frequency line  $\omega_a$  are uniformly bounded away from triple or higher resonances.

**Lemma 2.1.** If  $\mathbf{k}'' \in \mathbb{Z}^n \setminus \text{span}_{\mathbb{Z}} \{\mathbf{k}'\}$  is the second resonance of  $\lambda_a \left( a, \frac{P}{Q}, \frac{p}{q}, \hat{\omega}_{n-3}^* \right)^t$ ,

$$\langle \mathbf{k}'', \omega_a \rangle = 0, \quad |\mathbf{k}''| \leq K,$$

where  $\hat{\omega}_{n-3}^* \in \text{DC}(n-2, \alpha, \tau)$ , then for all

$$\mathbf{k} = (k_1, k_2, \dots, k_n) \in \mathbb{Z}^n \setminus \text{span}_{\mathbb{Z}} \{\mathbf{k}', \mathbf{k}''\}, \text{ and } |\mathbf{k}| \leq K,$$

we have the estimate

$$(2.4) \quad |\langle \mathbf{k}, \omega_a \rangle| \geq \frac{\alpha \cdot \inf_a \lambda_a}{2^\tau (qQ)^{\tau+1} (\|M'\|_\infty K)^{2\tau+1}}.$$

*Proof.* We use the linear transformation (2.3) to convert  $\omega_a$  to the vector

$$\omega'_a = M' \omega_a = \lambda_a \left( a, 0, \frac{d}{qQ}, \hat{\omega}_{n-3}^* \right)^t.$$

Denote by  $\tilde{\mathbf{k}}'' = (\tilde{k}_1'', \tilde{k}_2'', \dots, \tilde{k}_n'') := \mathbf{k}'' M'^{-1}$  so that we have

$$0 = \langle \mathbf{k}'', \omega_a \rangle = \langle \mathbf{k}'' M'^{-1}, M' \omega_a \rangle = \langle \tilde{\mathbf{k}}'', \omega'_a \rangle.$$

We have that  $\tilde{k}_1'' \neq 0$  since otherwise  $\langle \tilde{\mathbf{k}}'', \omega'_a \rangle = 0$  for all  $a$ , which is impossible considering that  $\hat{\omega}_{n-3}^*$  is Diophantine. We want to bound  $|\langle \mathbf{k}, \omega_a \rangle|$  from below for all

$$\mathbf{k} = (k_1, k_2, \dots, k_n) \in \mathbb{Z}^n \setminus \text{span}_{\mathbb{Z}} \{ \mathbf{k}', \mathbf{k}'' \}, \text{ and } |\mathbf{k}| \leq K.$$

We denote  $\tilde{\mathbf{k}} = (\tilde{k}_1, \tilde{k}_2, \dots, \tilde{k}_n) := \mathbf{k} M'^{-1}$  to get

$$\langle \mathbf{k}, \omega_a \rangle = \langle \mathbf{k} M'^{-1}, M' \omega_a \rangle = \langle \tilde{\mathbf{k}}, \omega'_a \rangle.$$

We introduce a new vector  $\bar{\mathbf{k}}$  as follows

$$\bar{\mathbf{k}} = \tilde{\mathbf{k}} - \frac{\tilde{k}_1''}{\tilde{k}_1''} \tilde{\mathbf{k}}'' = \frac{1}{\tilde{k}_1''} (\tilde{k}_1'' \tilde{\mathbf{k}} - \tilde{k}_1 \tilde{\mathbf{k}}'').$$

The new vector

$$\tilde{k}_1'' \bar{\mathbf{k}} = \tilde{k}_1'' \tilde{\mathbf{k}} - \tilde{k}_1 \tilde{\mathbf{k}}'' := (0, \bar{k}_2, \bar{k}_3, \hat{\mathbf{k}}_{n-3}) \in \mathbb{Z}^n$$

has zero first entry. We introduce further a new vector

$$\bar{\bar{\mathbf{k}}} = (0, d\bar{k}_2, d\bar{k}_3, qQ \hat{\mathbf{k}}_{n-3}) \in \mathbb{Z}^n.$$

We estimate the norm of  $\bar{\bar{\mathbf{k}}}$  as

$$|\bar{\bar{\mathbf{k}}}| \leq qQ |\tilde{k}_1'' \tilde{\mathbf{k}} - \tilde{k}_1 \tilde{\mathbf{k}}''| \leq 2qQ |\tilde{\mathbf{k}}''| \cdot |\tilde{\mathbf{k}}| \leq 2qQ (\|M'\|_\infty \cdot K)^2$$

Using the Diophantine conditions and the fact that  $\omega'_a$  has zero second entry, we have

$$(2.5) \quad \begin{aligned} |\langle \mathbf{k}, \omega_a \rangle| &= \left| \langle \tilde{\mathbf{k}}, \omega'_a \rangle \right| = \left| \langle \bar{\mathbf{k}}, \omega'_a \rangle \right| = \left| \frac{1}{\tilde{k}_1''} \langle (\tilde{k}_1'' \tilde{\mathbf{k}} - \tilde{k}_1 \tilde{\mathbf{k}}''), \omega'_a \rangle \right| \\ &= \frac{\lambda_a}{\tilde{k}_1'' qQ} \left| \langle \bar{\bar{\mathbf{k}}}, (0, 0, 1, \hat{\omega}_{n-3}^*) \rangle \right| \geq \frac{\inf_a \lambda_a}{\tilde{k}_1'' qQ} \frac{\alpha}{|\bar{\bar{\mathbf{k}}}|^\tau} \\ &\geq \frac{\alpha \inf_a \lambda_a}{2^\tau (qQ)^{\tau+1} (\|M'\|_\infty K)^{2\tau+1}}. \end{aligned}$$

□

Finally, we have the following fact.

**Lemma 2.2.** *Suppose the g.c.d. of the entries of  $\mathbf{k}''$  is 1. Then there exists a matrix  $M'' \in SL(n, \mathbb{Z})$  whose first row is  $\tilde{\mathbf{k}}'' = \mathbf{k}'' M'^{-1}$  with the first entry  $\tilde{k}_1'' \neq 0$ , and the second entry  $\tilde{k}_2'' = 0$ , such that  $M'' M' \omega_a = (0, 0, \hat{\omega}_{a, n-2}'')$ , where the  $\hat{\omega}_{a, n-2}''$  is a non resonant vector in  $\mathbb{R}^{n-2}$ . Moreover, the matrix  $M''$  has a special form such that*

$$\sigma(1, 2) M'' \sigma(1, 2) = \text{diag}\{1, M''_{n-1}\},$$

where  $M''_{n-1}$  is a matrix of order  $n-1$ , its first row is the vector  $\pi_{-2} \tilde{\mathbf{k}}''$ .

*Proof.* Denote by  $\omega'_a = M'\omega_a$ . We have  $\langle \mathbf{k}'', \omega_a \rangle = \langle \mathbf{k}'' M'^{-1}, \omega'_a \rangle = 0$ . We set the second entry of  $\tilde{\mathbf{k}}''$  to be zero and treat it as a vector in  $\mathbb{Z}^{n-1}$ . It is a basic fact in number theory that we can find  $n-2$  integer vectors which span unit volume together with  $\tilde{\mathbf{k}}''$ . By adding a number 0 as their second entry, we extend these vectors to be  $n$ -dimensional and put these vectors together to get an  $n \times n$  matrix  $M''$  whose first row is  $\tilde{\mathbf{k}}'' := \mathbf{k}''(M')^{-1}$ , the second row is  $(0, 1, 0, \dots, 0)$ , it satisfies the properties stated in the lemma.  $\square$

**2.2. Resonant submanifolds and their neighborhoods.** Let  $\omega_a, \mathbf{k}', \mathbf{k}''$  be as in Section 2.1. We define the single resonant sub-manifold associated to the vector  $\mathbf{k}'$

$$(2.6) \quad \Sigma' := \{y \mid \langle \mathbf{k}', \omega(y) \rangle = 0\}.$$

The following resonant curve  $\Gamma'_a$  determined by the frequency line  $\omega_a$  lies in  $\Sigma'$

$$(2.7) \quad \Gamma'_a = \partial h^{-1}(\omega_a) \subset \Sigma'.$$

In the single resonant sub-manifold we define the double resonant sub-manifold for the resonant vectors  $\mathbf{k}', \mathbf{k}''$

$$(2.8) \quad \Sigma'' := \{y \mid \langle \mathbf{k}', \omega(y) \rangle = \langle \mathbf{k}'', \omega(y) \rangle = 0\},$$

Next, we find a number  $\mu$  as the size of the neighborhood of the single resonant manifold such that within this neighborhood, either Proposition 2.1 or Proposition 2.2 (in the next subsection) applies. We use the notation  $B(a, r)$  to denote a ball of radius  $r$  centered at  $a$  and the notation  $B(A, r) := \cup_{a \in A} B(a, r)$  to denote the  $r$ -neighborhood of a set  $A$ .

We denote  $a''_1, a''_2, \dots, a''_m$  the list of points such that the corresponding frequency vector  $\omega_{a''}$  admits a second resonant vector  $\mathbf{k}''_{a''_i}$ ,  $i = 1, 2, \dots, m$ . The total number of such  $a''$ 's is bounded if we require  $|\mathbf{k}''| \leq K$  for some given large number  $K$ .

**Lemma 2.3.** *Let  $\omega_a, K, \mathbf{k}', \mathbf{k}''_{a''_i}$ ,  $i = 1, 2, \dots, m$  be as above. We define*

$$\mu = \frac{1}{2nK} \frac{\alpha \cdot \inf_a \lambda_a}{2^\tau (qQ)^{\tau+1} (\|M'\|_\infty K)^{2\tau+1}}.$$

*Let  $\mathbf{k}''_{a''_i}^\perp$  be the  $(n-1)$ -dimensional space orthogonal to the vector  $\mathbf{k}''_{a''_i}$ , then*

(1) *for all  $\omega$  in the neighborhood*

$$B(\omega_a, \mu) \setminus \bigcup_i B\left(\omega_{a''_i} + \mathbf{k}''_{a''_i}^\perp, \varepsilon^\sigma\right),$$

*and for sufficiently small  $\varepsilon$  we have*

$$|\langle \mathbf{k}, \omega \rangle| > \varepsilon^\sigma, \quad \forall \mathbf{k} \in \mathbb{Z}^n \setminus \text{span}_{\mathbb{Z}}\{\mathbf{k}'\}, \quad |\mathbf{k}| \leq K.$$

(2) *for all  $\omega$  in  $B(\omega_a, \mu) \cap B\left(\omega_{a''_i} + \mathbf{k}''_{a''_i}^\perp, \varepsilon^\sigma\right)$ , for each  $i$  and for all*

$$\mathbf{k} \in \mathbb{Z}^n \setminus \text{span}_{\mathbb{Z}}\left\{\mathbf{k}', \mathbf{k}''_{a''_i}\right\}, \quad \text{and } |\mathbf{k}| \leq K,$$

*we have*

$$(2.9) \quad |\langle \mathbf{k}, \omega \rangle| \geq \frac{\alpha \cdot \inf_a \lambda_a}{2^{\tau+1} (qQ)^{\tau+1} (\|M'\|_\infty K)^{2\tau+1}}.$$

*Proof.* Part (1). We consider two cases depending on if  $\mathbf{k}$  in the assumption is one of the double resonant vector  $\mathbf{k}_{a_i}''$  or not.

First we suppose  $\mathbf{k} = \mathbf{k}_{a_i}''$  for some  $i$ , then we get

$$|\langle \mathbf{k}, \omega \rangle| = |\langle \mathbf{k}, \omega_{a''} \rangle + \langle \mathbf{k}, \omega - \omega_{a''} \rangle| = |\langle \mathbf{k}, \omega - \omega_{a''} \rangle|.$$

By the assumption, the projection of  $\omega - \omega_{a''}$  to the vector  $\mathbf{k}_{a_i}''$  has length at least  $\varepsilon^\sigma$ . This completes the proof in the case  $\mathbf{k} = \mathbf{k}_{a_i}''$  for some  $i$  since  $|\mathbf{k}| \geq 1$ .

Next, suppose  $\mathbf{k} \neq \mathbf{k}_{a_i}''$ ,  $\forall i$ . Suppose there exists  $\omega^\dagger$  satisfying the assumption of part (1) but we have  $|\langle \mathbf{k}, \omega^\dagger \rangle| < \varepsilon^\sigma$ . We consider  $\omega_a$  in the frequency line segment that is within  $\mu$  distance of  $\omega^\dagger$ , i.e.  $|\omega^\dagger - \omega_a| \leq \mu$  for some  $a$ . First we must have the first entry  $k_1$  of  $\mathbf{k}$  is non vanishing. Otherwise, we have that the vector  $\mathbf{k}M'^{-1}$  has zero first entry and  $M'\omega_a = (a, 0, \frac{d_0}{qQ}, \hat{\omega}_{n-3}^*)$  has zero second entry. We have the estimate

$$(2.10) \quad |\langle \mathbf{k}, \omega_a \rangle| = |\langle \mathbf{k}M'^{-1}, M'\omega_a \rangle| \geq \frac{\alpha \cdot \inf_a \lambda_a}{(qQ)^{\tau+1}(\|M'\|_\infty K)^{2\tau+1}}$$

using the Diophantine property of  $\hat{\omega}_{n-3}^*$ . We get

$$(2.11) \quad \begin{aligned} |\langle \mathbf{k}, \omega^\dagger \rangle| &\geq |\langle \mathbf{k}, \omega_a \rangle| - |\langle \mathbf{k}, \omega^\dagger - \omega_a \rangle| \\ &\geq \frac{\alpha \cdot \inf_a \lambda_a}{(qQ)^{\tau+1}(\|M'\|_\infty K)^{2\tau+1}} - nK\mu \\ &\geq \frac{\alpha \cdot \inf_a \lambda_a}{2(qQ)^{\tau+1}(\|M'\|_\infty K)^{2\tau+1}} \gg \varepsilon^\sigma \end{aligned}$$

which is a contradiction. Next, since  $k_1 \neq 0$ , we change the first entry  $a$  of  $\omega_a$  to  $a'' := a - \frac{\langle \mathbf{k}, \omega_a \rangle}{k_1}$  to get another frequency vector  $\omega_{a''}$ . We have by definition  $\langle \mathbf{k}, \omega_{a''} \rangle = 0$ . This contradicts to the assumption that  $\mathbf{k} \neq \mathbf{k}_{a_i}''$ ,  $\forall i$ .

Part (2). For given  $\omega$  as assumed, we have  $|\omega - \omega_{a''}| \leq \mu$ . As  $\mathbf{k} \in \mathbb{Z}^n \setminus \text{span}_{\mathbb{Z}}\{\mathbf{k}', \mathbf{k}_{a_i}''\}$  and  $|\mathbf{k}| \leq K$ , we have the following estimate

$$\begin{aligned} |\langle \mathbf{k}, \omega \rangle| &= |\langle \mathbf{k}, \omega_{a''} \rangle + \langle \mathbf{k}, \omega - \omega_{a''} \rangle| \\ &\geq |\langle \mathbf{k}, \omega_{a''} \rangle| - |\langle \mathbf{k}, \omega - \omega_{a''} \rangle| \\ &\geq \frac{\alpha \cdot \inf_a \lambda_a}{2^\tau (qQ)^{\tau+1}(\|M'\|_\infty K)^{2\tau+1}} - nK\mu \\ &\geq \frac{\alpha \cdot \inf_a \lambda_a}{2^{\tau+1} (qQ)^{\tau+1}(\|M'\|_\infty K)^{2\tau+1}} \end{aligned}$$

where in the second inequality, we apply Lemma 2.1 and in the third inequality, we apply the definition of  $\mu$ .  $\square$

We choose  $\sigma < \frac{1}{r+2}$  such that  $\omega$  in the neighborhood of item (1) of Lemma 2.3 all satisfy the assumption of Proposition 2.1.

**2.3. Homogenization.** We first introduce the  $C^r$ -norm as follows.

**Definition 2.2.** For functions  $f(x, y)$  defined on a domain  $\mathcal{D} \times \mathbb{T}^n$ , we define the  $C^r$  norm as

$$|f|_r := \sup_{y \in \mathcal{D}} \left( \sum_{\mathbf{k} \in \mathbb{Z}^n} \sum_{|\alpha + \beta| \leq r} \left| \frac{\partial f^{\mathbf{k}}}{\partial y^\alpha} \right| (|k^\beta| + 1) \right)$$

where  $f^{\mathbf{k}}$  is the  $\mathbf{k}$ -th Fourier coefficient and we use the multi-index notation  $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ , etc. for  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ ,  $\beta = (\beta_1, \beta_2, \dots, \beta_n) \in \mathbb{Z}^n$ ,  $\alpha_i, \beta_i \geq 0$ ,  $i = 1, 2, \dots, n$ .

2.3.1. *Covering a  $\mu$ -neighborhood  $B(\omega_a, \mu)$  of the frequency line  $\omega_a$ .* For the frequency line segment  $\omega_a$ , consider any  $a''$  such that there is a second resonant vector  $\mathbf{k}_a''$ . The set  $\partial h^{-1}(B(\omega_{a''} + \mathbf{k}_a''^\perp, \varepsilon^\sigma))$  is an  $O(\varepsilon^\sigma)$ -neighborhood of  $\Sigma''$  in the space of action variables. We choose some large number  $\Lambda$  and cover this  $O(\varepsilon^\sigma)$ -neighborhood of  $\Sigma''$  using balls of radii  $\Lambda\varepsilon^\sigma$  centered on  $\Sigma''$ .

In the complement of the above  $O(\varepsilon^\sigma)$ -neighborhood of the double resonance sub-manifold  $\Sigma''$  inside  $B(\omega_a, \mu)$ , we use a covering by balls of radii  $\Lambda\sqrt{\varepsilon}$ . Let  $y^*$  to denote the center of any of the above balls. When we are considering a ball covering  $\Sigma''$ , we require  $y^* \in \Sigma''$ . Whenever a ball covers the single resonance sub-manifold  $\Sigma'$ , we require  $y^* \in \Sigma'$ .

2.3.2. *Homogenization.* We introduce the homogenization operator

$$(2.12) \quad \mathfrak{H}: \quad y - y^* := \sqrt{\varepsilon}Y, \quad t = \tau/\sqrt{\varepsilon}, \quad H(x, y) = \varepsilon\mathbf{H}(x, Y),$$

where  $Y, \tau, \mathbf{H}$  are the homogenized action variable, time and Hamiltonian respectively. The Hamiltonian becomes

$$(2.13) \quad \mathbf{H}(x, Y) = \frac{h(y^*)}{\varepsilon} + \frac{1}{\sqrt{\varepsilon}}\langle \omega^*, Y \rangle + \frac{1}{2}\langle \mathbf{A}Y, Y \rangle + \mathbf{V}(x) + \mathbf{P}(x, \sqrt{\varepsilon}Y),$$

where

- (1)  $\frac{h(y^*)}{\varepsilon} + \frac{1}{\sqrt{\varepsilon}}\langle \omega^*, Y \rangle + \frac{1}{2}\langle \mathbf{A}Y, Y \rangle$  is the first three terms of the Taylor expansion of  $h(y)$  around  $y^*$ ;
- (2)  $\omega^* = \frac{\partial h}{\partial y}(y^*)$ ;
- (3)  $\mathbf{A} = \frac{\partial^2 h}{\partial y^2}(y^*)$  is a positive definite constant matrix;
- (4)  $\mathbf{V}(x) = P(x, y^*)$  is the constant term in the Taylor expansion of  $P(x, y)$  with respect to the variable  $y$ ;
- (5) The term  $\mathbf{P}$  has a decomposition  $\mathbf{P} = \mathbf{P}_I + \mathbf{P}_{II}$  where

$$(2.14) \quad \begin{aligned} \mathbf{P}_I &= \frac{1}{\varepsilon} \left( h(y^* + \sqrt{\varepsilon}Y) - h(y^*) - \sqrt{\varepsilon}\langle \omega, Y \rangle - \frac{\varepsilon}{2}\langle \mathbf{A}Y, Y \rangle \right), \\ &= \frac{\sqrt{\varepsilon}}{6} \sum_{1 \leq i, j, k \leq n} Y_i Y_j Y_k \int_0^1 \frac{\partial^3 h}{\partial y_i \partial y_j \partial y_k} (ty + (1-t)y^*) t^2 dt \\ \mathbf{P}_{II} &= P(x, y^* + \sqrt{\varepsilon}Y) - P(x, y^*) \\ &= \sqrt{\varepsilon} \left\langle Y, \int_0^1 \frac{\partial P}{\partial y}(x, ty + (1-t)y^*) dt \right\rangle \end{aligned}$$

The estimate of the remainders  $\mathbf{P}_I$  and  $\mathbf{P}_{II}$  relies on the size of the balls. When a ball of radius  $\varepsilon^\sigma$  is used to cover  $\Sigma''$ , we have the following estimate.

$$(2.15) \quad \left| \frac{\partial^{|\alpha|+|\beta|} \mathbf{P}_{II}}{\partial x^\alpha \partial Y^\beta} \right| \leq C_r \sqrt{\varepsilon}^{|\beta|}, \quad |\beta| > 0, \text{ for } |Y| \leq \varepsilon^{\sigma-1/2},$$

where  $C_r > 0$  is an upper bound of  $C^{r+1}$ -norm of  $h$  and  $P$ , in terms of  $(x, y)$ .  $\mathbf{P}_I$  is independent of  $x$  and is large for large  $Y$ , however we have

$$(2.16) \quad \left| \frac{\partial^\beta \mathbf{P}_I}{\partial Y^\beta} \right| = O(\sqrt{\varepsilon}) \text{ for } |Y| < \text{const. as } \varepsilon \rightarrow 0 \text{ and } |\beta| \leq r.$$

**2.4. The normal form around single resonance.** In this section, we work out it.

**Proposition 2.1.** *Given any small  $\delta$ , suppose  $y^*$  is such that  $\omega^* = \partial h(y^*) \in B(\omega_a, \mu)$  and satisfies the following for some constant  $C_1 > 0$ ,*

$$(2.17) \quad |\langle \mathbf{k}, \omega^* \rangle| > C_1 \varepsilon^{\frac{1}{r+2}} \delta^{-\frac{1}{2}}, \quad \mathbf{k} \in \mathbb{Z}^n \setminus \text{span}_{\mathbb{Z}}\{\mathbf{k}'\}, \quad |\mathbf{k}| \leq K = (\delta/3)^{-\frac{1}{2}},$$

Denote the frequency map by  $\omega(y) = \partial h(y)$ . Then there exists a symplectic transformation  $\phi$  defined on  $(\varepsilon^{-1/2} B(y^*, \sqrt{\varepsilon})) \times \mathbb{T}^n$ , which is  $o_{\varepsilon \rightarrow 0}(1)$  close to identity in the  $C^{r-2}$  topology,  $|\phi - \text{id}|_{r-2} = O(\varepsilon^{\frac{r}{2(r+2)}} \delta^{\frac{1}{2}})$ , sending the Hamiltonian  $H \in C^r$  of (2.13) to the following form provided  $\varepsilon$  is sufficiently small

$$(2.18) \quad \begin{aligned} H \circ \phi(x, Y) = & \frac{1}{\varepsilon} h(y^*) + \frac{1}{\sqrt{\varepsilon}} \langle \omega^*, Y \rangle + \frac{1}{2} \langle AY, Y \rangle + V(\langle \mathbf{k}', x \rangle) \\ & + \delta R_I(x) + \varepsilon^{\frac{r}{2(r+2)}} R_{II}(x, Y), \end{aligned}$$

where  $V$  consists of all the Fourier modes in  $\mathbf{V}$  with  $\mathbf{k} \in \text{span}_{\mathbb{Z}}\{\mathbf{k}'\}$ , the remainder  $\delta R_I$  consists of all the Fourier modes in  $\mathbf{V}$  with  $|\mathbf{k}| > K$ , and the  $C^{r-2}$  norm of  $V, R_I, R_{II}$  satisfies  $|V|_r, |R_I|_{r-2}, |R_{II}|_{r-2} \leq O(1)$  as  $\delta, \varepsilon \rightarrow 0$ .

**Remark 2.1.** *Considering the part (1) of Lemma 2.3, the assumption (2.17) on the frequency  $\omega^*$  is satisfied if  $y^*$  is not the center of the balls covering  $\varepsilon^\sigma$ -neighborhood of  $\Sigma''$  ( $\sigma < \frac{1}{r+2}$ ). By the choice of the covering, such a ball centered at  $y^*$  has radius  $\sqrt{\varepsilon}$ .*

*Proof.* We decompose the Hamiltonian (2.13) as follows

$$H = \frac{1}{\varepsilon} h(y^*) + \frac{1}{\sqrt{\varepsilon}} \langle \omega^*, Y \rangle + \frac{1}{2} \langle AY, Y \rangle + V(\langle \mathbf{k}', x \rangle) + R_{\leq}(x) + R_{>}(x) + P(x, \sqrt{\varepsilon} Y).$$

where  $V(\langle \mathbf{k}', x \rangle)$  contains all the Fourier modes of  $\mathbf{V}(x)$  in the  $\text{span}_{\mathbb{Z}}\{\mathbf{k}'\}$ , the remaining term  $R_{\leq}(x) + R_{>}(x)$  contains all the Fourier modes of  $\mathbf{V}(x)$  in  $\mathbb{Z}^n \setminus \text{span}_{\mathbb{Z}}\{\mathbf{k}'\}$ . Given  $\delta$ , let  $K = \delta^{-1/2}$ . The Fourier modes for  $\mathbf{k} \in \mathbb{Z}^n \setminus \text{span}_{\mathbb{Z}}\{\mathbf{k}'\}$  with  $|\mathbf{k}| \leq K$  are put in  $R_{\leq}$  and those with  $|\mathbf{k}| > K$  are put in  $R_{>}$ . We have the estimate  $|R_{>}|_{r-2} \leq \delta$ .

Only one step of KAM iteration is needed. We use a new Hamiltonian  $\sqrt{\varepsilon} F$  whose induced time-1 map  $\phi_{\sqrt{\varepsilon} F}^1$  gives rise to a symplectic transformation

$$\begin{aligned} H \circ \phi_{\sqrt{\varepsilon} F}^1 = & H + \sqrt{\varepsilon} \{H, F\} + \frac{\varepsilon}{2} \int_0^1 (1-t) \{ \{H, F\}, F \} (\Phi_{\sqrt{\varepsilon} F}^t) dt \\ = & \frac{h(y^*)}{\varepsilon} + \frac{1}{\sqrt{\varepsilon}} \langle \omega^*, Y \rangle + \frac{1}{2} \langle AY, Y \rangle + V(\langle \mathbf{k}', x \rangle) + \left\langle \omega^*, \frac{\partial F}{\partial x} \right\rangle \\ & + R_{\leq}(x) + R_{>}(x) + P(x, \sqrt{\varepsilon} Y) \\ & + \sqrt{\varepsilon} \left\langle AY + \frac{\partial P}{\partial Y}, \frac{\partial F}{\partial x} \right\rangle + \frac{\varepsilon}{2} \int_0^1 (1-t) \{ \{H, F\}, F \} (\Phi_{\sqrt{\varepsilon} F}^t) dt, \end{aligned}$$

where  $F$  solves the cohomological equation

$$R_{\leq}(x) + \left\langle \omega^*, \frac{\partial F}{\partial x} \right\rangle = 0.$$

Notice that  $F$  is a function of only  $x$ . Notice also  $|P|_r = 1$  and  $\mathbf{V}(x) = P(x, y^*)$ , so we get  $|F|_r = O(\varepsilon^{-\frac{1}{r+2}} \delta^{\frac{1}{2}})$  as  $\varepsilon, \delta \rightarrow 0$  by solving the cohomological equation under the assumption (2.17).

Let  $\delta R_I = R_{>}$ . As we choose  $K = \delta^{-1/2}$  and  $V \in C^r$ , we have  $|R_I|_{r-2} \leq O(1)$ . Let

$$\begin{aligned} \varepsilon^{\frac{r}{2(r+2)}} R_{II} &= P(x, \sqrt{\varepsilon}Y) + \sqrt{\varepsilon} \left\langle AY + \frac{\partial P}{\partial Y}, \frac{\partial F}{\partial x} \right\rangle \\ &\quad + \frac{\varepsilon}{2} \int_0^1 (1-t) \{ \{H, F\}, F \} \left( \phi_{\sqrt{\varepsilon}F}^t \right) dt. \end{aligned}$$

We have

- (1)  $|P|_{r+1} \leq |P_I|_{r-1} + |P_{II}|_{r-1} \leq C_r \varepsilon^{1/2}$  from formula (2.15) and (2.16),
- (2) using the derivative estimates of  $F$  and the fact that  $Y = O(1)$  we find

$$\left| \sqrt{\varepsilon} \left\langle AY + \frac{\partial P}{\partial Y}, \frac{\partial F}{\partial x} \right\rangle \right|_{r-1} = O(\delta^{1/2} \varepsilon^{1/2 - \frac{1}{r+2}}).$$

- (3) since  $\{H, F\} = \left\{ \frac{1}{\sqrt{\varepsilon}} \langle \omega^*, Y \rangle + \frac{1}{2} \langle AY, Y \rangle + V(x) + P, F \right\}$ , we find

$$\left| \frac{\varepsilon}{2} \int_0^1 (1-t) \{ \{H, F\}, F \} \left( \phi_{\sqrt{\varepsilon}F}^t \right) dt \right|_{r-2} = O(\delta^{1/2} \varepsilon^{1/2 - \frac{1}{r+2}}).$$

Therefore, we have  $|R_{II}|_{r-2} \leq O(\delta^{1/2})$  and can make the term  $\varepsilon^{\frac{r}{2(r+2)}} R_{II}$  arbitrarily small by decreasing  $\varepsilon$ . The proof is now complete.  $\square$

**2.5. The normal form around double resonance.** In this section, we obtain the normal form of the Hamiltonian system in a  $\varepsilon^\sigma$  neighborhood of  $\Sigma''$ .

**Proposition 2.2.** *For all small  $\delta > 0$ , there exists a symplectic transformation  $\phi$  defined on  $\varepsilon^{-1/2} \partial h^{-1}(B(\omega^*, \varepsilon^\sigma)) \times \mathbb{T}^n$ , which is  $O(\sqrt{\varepsilon} \delta^{-(\tau+1)})$  close to identity in the  $C^{r-2}$  topology,  $|\phi - \text{id}|_{r-2} = O(\sqrt{\varepsilon} \delta^{-(\tau+1)})$ , sending the Hamiltonian  $H \in C^r$  of (2.13) to the following form provided  $\varepsilon$  is sufficiently small*

$$\begin{aligned} (2.19) \quad H \circ \phi(x, Y) &= \frac{1}{\varepsilon} h(y^*) + \frac{1}{\sqrt{\varepsilon}} \langle \omega^*, Y \rangle + \frac{1}{2} \langle AY, Y \rangle + P_I(Y) \\ &\quad + V(\langle \mathbf{k}', x \rangle, \langle \mathbf{k}'', x \rangle) + \delta R_I(x) + \varepsilon^\sigma R_{II}(x, Y), \end{aligned}$$

where  $V$  consists of all the Fourier modes in  $V$  with  $\mathbf{k} \in \text{span}_{\mathbb{Z}}\{\mathbf{k}', \mathbf{k}''\}$ , the remainder  $\delta R_I$  consists of all the Fourier modes in  $V$  with  $|\mathbf{k}| > K$ , and the  $C^{r-2}$  norm of  $V, R_I, R_{II}$  satisfies  $|V|_r, |R_I|_{r-2}, |R_{II}|_{r-2} \leq O(1)$  as  $\varepsilon \rightarrow 0$ .

*Proof.* We decompose the Hamiltonian (2.13) as follows

$$\begin{aligned} H &= \frac{1}{\varepsilon} h(y^*) + \frac{1}{\sqrt{\varepsilon}} \langle \omega^*, Y \rangle + \frac{1}{2} \langle AY, Y \rangle + V(\langle \mathbf{k}', x \rangle, \langle \mathbf{k}'', x \rangle) \\ &\quad + R_{\leq}(x) + R_{>}(x) + P(x, \sqrt{\varepsilon}Y). \end{aligned}$$

where

- (1)  $V(\langle \mathbf{k}', x \rangle, \langle \mathbf{k}'', x \rangle)$  contains all the Fourier modes of  $V(x)$  in the  $\text{span}_{\mathbb{Z}}\{\mathbf{k}', \mathbf{k}''\}$ ;
- (2) the remaining term  $R_{\leq}(x) + R_{>}(x)$  contains all the Fourier modes of  $V(x)$  in  $\mathbb{Z}^n \setminus \text{span}_{\mathbb{Z}}\{\mathbf{k}', \mathbf{k}''\}$ . Given  $\delta$ , we choose  $K$  satisfying  $K = \delta^{-1/2}$ . The Fourier modes for  $\mathbf{k} \in \mathbb{Z}^n \setminus \text{span}_{\mathbb{Z}}\{\mathbf{k}', \mathbf{k}''\}$  with  $|\mathbf{k}| \leq K$  are put in  $R_{\leq}$  and those with  $|\mathbf{k}| > K$  are put in  $R_{>}$ . We have the estimate  $|R_{>}|_{C^{r-2}} \leq \delta$ .



To run the KAM machine for one step, we use a new Hamiltonian  $\sqrt{\varepsilon}F$  whose induced time-1 map  $\phi_{\sqrt{\varepsilon}F}^1$  gives rise to a symplectic transformation

$$\begin{aligned} \mathbf{H} \circ \phi_{\sqrt{\varepsilon}F}^1 &= \mathbf{H} + \sqrt{\varepsilon}\{\mathbf{H}, F\} + \frac{\varepsilon}{2} \int_0^1 (1-t)\{\{\mathbf{H}, F\}, F\}(\Phi_{\sqrt{\varepsilon}F}^t) dt \\ &= \frac{h(y^*)}{\varepsilon} + \frac{1}{\sqrt{\varepsilon}}\langle \omega^*, Y \rangle + \frac{1}{2}\langle \mathbf{A}Y, Y \rangle + V(\langle \mathbf{k}', x \rangle, \langle \mathbf{k}'', x \rangle) \\ &\quad + \left\langle \omega^*, \frac{\partial F}{\partial x} \right\rangle + R_{\leq}(x) + R_{>}(x) + \mathbf{P}(x, \sqrt{\varepsilon}Y) \\ &\quad + \sqrt{\varepsilon}\left\langle \mathbf{A}Y + \frac{\partial \mathbf{P}}{\partial Y}, \frac{\partial F}{\partial x} \right\rangle \\ &\quad + \frac{\varepsilon}{2} \int_0^1 (1-t)\{\{\mathbf{H}, F\}, F\}(\Phi_{\sqrt{\varepsilon}F}^t) dt, \end{aligned}$$

where  $F$  solves the cohomological equation

$$R_{\leq}(x) + \left\langle \omega^*, \frac{\partial F}{\partial x} \right\rangle = 0.$$

Formula (2.9) in Lemma 2.3 is used to bound  $|\langle \mathbf{k}, \omega^* \rangle|$  from below for  $|\mathbf{k}| \leq K$  by  $CK^{-(2\tau+1)}$  for some constant  $C$  independent of  $\delta, \varepsilon$ . Notice that  $F$  is a function of only  $x$ . Notice  $|P|_r = 1$  and  $V(x) = P(x, y^*)$ , so we get  $|F|_r = O(K^{(2\tau+1)})$  as  $\varepsilon, \delta \rightarrow 0$  by solving the cohomological equation.

Let  $\delta R_I = R_{>}$ . As we choose  $K = \delta^{-1/2}$  and  $V \in C^r$ , we have  $|R_I|_{r-2} \leq O(1)$ . Let

$$\begin{aligned} \varepsilon^\sigma R_{II} &= \mathbf{P}_{II}(x, \sqrt{\varepsilon}Y) + \sqrt{\varepsilon}\left\langle \mathbf{A}Y + \frac{\partial \mathbf{P}}{\partial Y}, \frac{\partial F}{\partial x} \right\rangle \\ &\quad + \frac{\varepsilon}{2} \int_0^1 (1-t)\{\{\mathbf{H}, F\}, F\}(\phi_{\sqrt{\varepsilon}F}^t) dt. \end{aligned}$$

We have

- (1) we obtain from formula (2.15) that  $|\mathbf{P}_{II}|_{r-1} \leq C_r \varepsilon^\sigma$ ,
- (2) using the derivative estimates of  $F$  and the fact that  $Y = O(\varepsilon^{\sigma-1/2})$  we find

$$\left| \sqrt{\varepsilon}\left\langle \mathbf{A}Y + \frac{\partial \mathbf{P}}{\partial Y}, \frac{\partial F}{\partial x} \right\rangle \right|_{r-1} = O(\varepsilon^\sigma K^{(2\tau+1)}).$$

- (3) since  $\{\mathbf{H}, F\} = \left\{ \frac{1}{\sqrt{\varepsilon}}\langle \omega^*, Y \rangle + \frac{1}{2}\langle \mathbf{A}Y, Y \rangle + V(x) + \mathbf{P}, F \right\}$ , we find

$$\left| \frac{\varepsilon}{2} \int_0^1 (1-t)\{\{\mathbf{H}, F\}, F\}(\phi_{\sqrt{\varepsilon}F}^t) dt \right|_{r-2} = O\left(\varepsilon^\sigma K^{2(2\tau+1)}\right).$$

Therefore, we have  $|R_{II}|_{r-2} \leq O(K^{2(2\tau+1)})$  and can make the term  $\varepsilon^\sigma R_{II}$  arbitrarily small by decreasing  $\varepsilon$ . The proof is now complete.  $\square$

### 3. THE CHOICE OF DIFFUSION PATH: FROM SINGLE, DOUBLE TO COMPLETE RESONANCES

**3.1. Normally hyperbolic invariant manifold for Hamiltonian system.** In this section, we introduce the theory of normally hyperbolic invariant manifold (NHIM). Besides its persistence under perturbation, we shall use the fact: a Hamiltonian system restricted on a NHIM is still a Hamiltonian system with less degrees of freedom. It was discovered and used intensively by Delshams, de la Llave and Seara in their geometric

theory of Arnold diffusion of a *a priori* unstable type to get a Hamiltonian restricted on a NHIC whose time-periodic map is the *inner map* (cf. [DLS08]).

We introduce the definition of the normally hyperbolic invariant manifold following [DLS00, DLS08]. As remarked by de la Llave, the definition in [DLS00, DLS08] is less general than that of [Fe] but better adapted to the symplectic structure.

**Definition 3.1.** *Let  $f : M \rightarrow M$  be a  $C^r$ -diffeomorphism on a smooth manifold  $M$  with  $r > 1$ . Let  $\Pi \subset M$  be a submanifold invariant under  $f$ ,  $f(\Pi) = \Pi$ . We say that  $\Pi$  is a normally hyperbolic invariant manifold if there exist a constant  $C > 0$ , rates  $0 < \lambda < \mu^{-1} < 1$  and a splitting for every  $x \in \Pi$*

$$T_x M = E_x^s \oplus E_x^u \oplus T_x \Pi$$

in such a way that

$$\begin{aligned} v \in E_x^s &\Leftrightarrow |Df^k(x)v| \leq C\lambda^k|v|, \quad k \geq 0, \\ v \in E_x^u &\Leftrightarrow |Df^k(x)v| \leq C\lambda^{|k|}|v|, \quad k \leq 0, \\ v \in T_x \Pi &\Leftrightarrow |Df^k(x)v| \leq C\mu^k|v|, \quad k \in \mathbb{Z}. \end{aligned}$$

**Theorem 3.1** (Theorem A.14 of [DLS00]). *Let  $\Pi_X \subset M$  - not necessarily compact - be hyperbolic for the map  $f_X$  generated by the vector field  $X$ , which is uniformly  $C^r$  in a neighborhood  $U$  of  $\Pi_X$  such that  $\text{dist}(M \setminus U, \Pi_X) > 0$ . Let  $f_Y$  be the  $C^r$ -map generated by another vector field  $Y$  which is sufficiently close to  $X$  in the  $C^1$ -topology. Then, we can find a manifold  $\Pi_Y$  which is hyperbolic for  $Y$  and close to  $\Pi_X$  in the  $C^{\min\{r, \lambda/\mu-\delta\}}$  topology. The Lyapunov exponents for  $\Pi_Y$  are arbitrarily close to those of  $\Pi_X$  if  $Y$  is sufficiently close to  $X$  in the  $C^1$  topology. The manifold  $\Pi_Y$  is the only  $C^1$  manifold close to  $\Pi_X$  in the  $C^0$  topology, and invariant under the flow of  $Y$ .*

When the normally hyperbolic flow is Hamiltonian, we have the following theorem saying that the restriction of the Hamiltonian system to the central manifold is also Hamiltonian with less number of degrees of freedom.

**Theorem 3.2** (Theorem 23 and 26 of [DLS08]). *Suppose  $M$  is endowed with a (an exact) symplectic form  $\omega$ . Let  $f_\varepsilon : M \rightarrow M$  be a  $C^r$  family of Hamiltonomorphisms,  $r \geq 2$  preserving  $\omega$ . Assume that  $\Pi \subset M$  is a normally hyperbolic invariant manifold for  $f_0$  with rate  $\lambda, \mu$ .*

- (1) *Then for sufficiently small  $\varepsilon$ , there exist  $C^\ell$ -families of diffeomorphisms  $k_\varepsilon$ ,  $r_\varepsilon$  with  $\ell \leq \min \left\{ r, \left\lfloor \frac{\ln \lambda}{\ln \mu} \right\rfloor \right\}$ , satisfying  $f_\varepsilon \circ k_\varepsilon = k_\varepsilon \circ r_\varepsilon$  where  $k_\varepsilon$  is the map such that  $k_\varepsilon(\Pi) = \Pi_\varepsilon$  and  $r_\varepsilon : \Pi \rightarrow \Pi$  is the restricted map on  $\Pi$ .*
- (2) *We denote by  $\mathcal{R}_\varepsilon$  the generating vector field corresponding to  $r_\varepsilon$  defined by  $\frac{d}{d\varepsilon} r_\varepsilon = \mathcal{R}_\varepsilon \circ r_\varepsilon$ . Then we have*
  - $k_\varepsilon^* \omega = \omega_\Pi$  is a (an exact) symplectic form on  $\Pi$ . It is independent of  $\varepsilon$ .
  - The vector field  $\mathcal{R}_\varepsilon$  is (exactly) Hamiltonian vector field with respect to  $\omega_\Pi$ . Moreover, its (global) Hamiltonian is  $R_\varepsilon = F_\varepsilon \circ k_\varepsilon$  where  $F_\varepsilon$  is the Hamiltonian for  $f_\varepsilon$ .

**3.2. Normally hyperbolic invariant cylinder (NHIC) in the regime of single resonances.** Using Formula (2.3), we introduce a linear symplectic transformation denoted by  $\mathfrak{M}'$ ,

$$x' = M'x, \quad Y' = (M')^{-t}Y.$$

The frequency  $\omega_a$  is transformed to  $\omega'_a = M'\omega_a = \lambda_a(a, 0, \frac{d}{qQ}, \hat{\omega}_{n-2}^*)$ .

In Proposition 2.1, we choose  $y^* \in \Sigma'$  such that  $\omega' = M'\omega^*$  has zero as the second entry. Applying the symplectic transformation  $\mathfrak{M}'$  to the normal form (2.18), we get the following system

$$\begin{aligned} \mathfrak{M}'^{-1*} \mathbf{H} \circ \phi = & \frac{1}{\varepsilon} h(y^*) + \frac{1}{\sqrt{\varepsilon}} \langle \omega'^*, Y' \rangle + \frac{1}{2} \langle AY', Y' \rangle + V(x'_2) \\ & + \delta R_I(x) + \varepsilon^{\frac{r}{2(r+2)}} R_{II}(x, Y), \end{aligned}$$

where  $A = M'AM'^t$  and  $R_I(x') = R_I(M'^{-1}x)$ ,  $R_{II}(x', Y') = \mathfrak{M}'^{-1*} R_{II}(x, Y)$ .

We denote by  $Z'(x', y') = \mathfrak{M}'^{-1*} Z(\langle \mathbf{k}', x \rangle, y)$  where  $Z(\langle \mathbf{k}', x \rangle, y)$  consists of all the Fourier modes of  $P(x, y)$  in the  $\text{span}_{\mathbb{Z}}\{\mathbf{k}'\}$ . From the definition of  $V(\langle \mathbf{k}', x \rangle)$  in Proposition 2.1, we get that  $V(\langle \mathbf{k}', x \rangle) = Z(\langle \mathbf{k}', x \rangle, y^*)$ .

The following assumption is shown to be generic ([CZ1] and Theorem B.1 of [C12]):

**(H1):** For each  $y \in \Gamma'_a$ , the function  $Z'$  is non-degenerate at its globally maximal point, i.e.  $-\partial_{x'_2}^2 Z'(x'_2, y) > 0$  holds if  $x'_2$  is a globally maximal point. The number of global maximum is unique except for finitely many  $y$ 's there are two global maximums.

Under this genericity hypothesis, we get that the function  $V(x'_2)$  has non degenerate global maximum point for all  $y^*$  in a  $\mu$ -neighborhood of  $\Gamma'_a$  and  $\delta$  small enough, since  $\mu = o_{\delta \rightarrow 0}(1)$ .

In (2.18), we neglect the remainder  $\delta R_I + \varepsilon^{\frac{r}{2(r+2)}} R_{II}(x, Y)$  to get that the remaining system

$$\mathbf{H}'_0 := \frac{1}{\varepsilon} h(y^*) + \frac{1}{\sqrt{\varepsilon}} \langle \omega'^*, Y' \rangle + \frac{1}{2} \langle AY', Y' \rangle + V(x'_2)$$

admits a NHIC given by

$$(3.1) \quad \Pi_{n-1}^0 : \left\{ \dot{Y}'_2 = \frac{\partial V}{\partial x'_2} = 0, \quad \dot{x}'_2 = \frac{1}{2} \frac{\partial \langle AY', Y' \rangle}{\partial Y'_2} = \sum_{i=1}^n A_{2i} Y'_i = 0 \right\}.$$

The normal Lyapunov exponent does not depend on  $\varepsilon$  or  $\delta$ . Restricted to the NHIC we get a system with one less degrees of freedom due to Theorem 3.2. We denote by

$$\bar{Y} = \pi_{-2} Y', \quad \bar{x} = \pi_{-2} x', \quad \bar{\omega}^* = \pi_{-2} \omega'^*,$$

and by  $\langle \bar{A}\bar{Y}, \bar{Y} \rangle$  the quadratic form  $\langle AY', Y' \rangle$  obtained by substituting  $Y'_2$  solved from (3.1). The restricted system now has the form

$$\bar{\mathbf{H}}_0 = \frac{1}{\sqrt{\varepsilon}} \langle \bar{\omega}^*, \bar{Y} \rangle + \frac{1}{2} \langle \bar{A}\bar{Y}, \bar{Y} \rangle$$

where we have dropped some constants for simplification. Also, it can be verified directly with the help of (3.1) that  $\bar{\mathbf{H}}_0$  gives rise to the Hamiltonian flow restricted to the NHIC.

The system  $\mathbf{H}$  (without the linear transformation) is defined in a  $\Lambda$  ball in the  $Y$  variables since the homogenization is done in a  $\Lambda\sqrt{\varepsilon}$  ball. To the locally finite covering that we use before homogenization, we associate a partition of unity  $\sum \psi_i(Y) = 1$  to the homogenized system, where  $\psi_i(Y) \in C^\infty$ ,  $\psi_i(Y) = 1$  if  $|Y| < (1-d)\Lambda$  and is zero for  $|Y| > (1-0.5d)\Lambda$  where  $2d$  is smaller than the Lebesgue number of the covering. To apply the NHIM theorem, we replace the remainder  $\delta R_I + \varepsilon^{\frac{r}{2(r+2)}} R_{II}$  in (2.18) by  $\psi_i(M''Y)(\delta R_I + \varepsilon^{\frac{r}{2(r+2)}} R_{II})$ .

We next apply the NHIM theorems 3.1 and 3.2. In the Hamiltonian equations, the vector field in the center is fast  $\dot{x} = \frac{\bar{\omega}^*}{\sqrt{\varepsilon}} + O(1)$ . However, the NHIM theorems are

still applicable since the large term  $\frac{\bar{\omega}^*}{\sqrt{\varepsilon}}$  is constant, which does not contribute to the derivatives of the Hamiltonian flow, since the derivatives of the Hamiltonian flow are solutions of the variational equations where  $\frac{\bar{\omega}^*}{\sqrt{\varepsilon}}$  does not appear. The time-one map of the Hamiltonian flow admits NHIM given by (3.1) satisfying Definition 3.1 where none of the constants  $\lambda, \mu, C$  depends on  $\varepsilon$ . The perturbation to the Hamiltonian is  $C^{r-2}$  small hence is a  $C^{r-3}$  small perturbation to the vector field. Only the  $C^1$  smallness of the perturbation to the vector fields, the normal hyperbolicity and the derivatives of the time- $T$  map for some  $T$  (can be chosen to be 1 in our case) are needed in the proof of the NHIM theorem [Fe]. Applying the NHIM theorems we get a NHIC which is  $\delta$  close to the unperturbed one in the  $C^{r-2}$ -topology as the center Lyapunov exponents are zero. The NHIC for the modified system agrees with that for the system 2.18 in the region where  $\psi_i = 1$  and in the overlapping region of two balls in the covering the NHICs coincide due to the local uniqueness. In this case, we apply Theorem 3.2 to restrict the system to the NHIC to get a Hamiltonian system with one degree of freedom less

$$(3.2) \quad \bar{H}_\delta = \frac{1}{\sqrt{\varepsilon}} \langle \bar{\omega}^*, \bar{Y} \rangle + \frac{1}{2} \langle \bar{A}\bar{Y}, \bar{Y} \rangle + \delta \bar{R}_I(\bar{x}) + \varepsilon^{\frac{r}{2(r+2)}} \bar{R}_{II}(\bar{x}, \bar{Y}),$$

where  $|\bar{R}_i - R_i|_{(3.1), r-2} = O_{\delta \rightarrow 0}(\delta^2)$ ,  $i = I, II$ . To obtain this form of the Hamiltonian, we write the perturbed NHIC as a graph over the unperturbed cylinder

$$Y'_2 = Y_2'^* + \delta G_Y(\bar{Y}, \bar{x}), \quad x'_2 = x_2'^* + \delta G_x(\bar{Y}, \bar{x})$$

for some  $G_x, G_Y \in C^{r-2}$ , the  $\delta = 0$  case corresponding to the unperturbed NHIC. Substituting the graph  $Y'_2, x'_2$  to the system  $\mathfrak{M}^{r-1*} \mathbf{H} \circ \phi(x, Y)$  above, we get a  $\delta^2$  correction to the restriction of  $R_i$ ,  $i = I, II$  on the NHIC.

**3.3. Strong double resonances and shear transformation.** The number of double resonance treated in Proposition 2.2 depends on  $\delta$ . However, most of the double resonances are weak and can be treated as single resonance. The number of strong double resonances is independent of  $\delta, \varepsilon$ .

**3.3.1. Criteria of distinguishing weak and strong double resonance and the reduction of order at weak double resonance.** Consider a double resonance associated to the vector  $\mathbf{k}''$ . We decompose  $Z$  in (2.19) in Proposition 2.2 as

$$(3.3) \quad Z(\langle \mathbf{k}', x \rangle, \langle \mathbf{k}'', x \rangle) = Z'(\langle \mathbf{k}', x \rangle) + Z''(\langle \mathbf{k}', x \rangle, \langle \mathbf{k}'', x \rangle)$$

where  $Z'$  includes all the Fourier harmonics in the  $\text{span}\{\mathbf{k}'\}$  and  $Z''$  contains the rest.

Notice  $Z''$  must contain at least one term with  $\mathbf{k}''$ . Since  $\mathbf{k}'$  does not depend on  $\delta$ , we get  $|Z''|_{C^{r-2}} \leq \frac{C}{|\mathbf{k}''|^{r/2}}$  for some constant  $C$  independent of  $\delta$ . We first treat  $Z'' + \delta R$  as a perturbation to the truncated Hamiltonian  $\frac{1}{\sqrt{\varepsilon}} \langle \omega^*, Y \rangle + \frac{1}{2} \langle AY, Y \rangle + Z'(\langle \mathbf{k}', x \rangle)$ , which has a NHIC following from exactly the same reasoning as the previous section. There is a threshold denoted by  $\delta$  that is the maximal allowable  $C^1$  norm of the perturbation for applying the NHIM Theorem 3.1 to the NHIC in the truncated Hamiltonian. The threshold  $\delta$  does not depend on  $\delta, \varepsilon$  so we get when  $\delta > 2 \frac{C}{|\mathbf{k}''|^{r/2}}$ , we treat the corresponding double resonance point as a single resonance, otherwise we treat the point as a strong double resonance point. The total number of strong double resonance points are bounded by  $\left(\frac{2C}{\delta}\right)^{n/r'}$  which is independent of  $\varepsilon, \delta$  for given generic  $P$ .

After the reduction of order and also undoing the homogenization, we obtain a Hamiltonian system of the same form as (3.2) with  $\delta$  replaced by  $\delta$ .

3.3.2. *The shear transformation for the strong double resonances.* Next, we write the homogenized Hamiltonian in the  $\varepsilon^{\sigma-1/2}$  neighborhood of the double resonant submanifold  $\Sigma''$  into a standard form. We denote by  $a''$  the  $a$  for which there is a second resonance vector  $\mathbf{k}''$ .

In the next lemma, we are going to introduce a linear symplectic transformation that is in  $SL(2n, \mathbb{R})$  but not in  $SL(2n, \mathbb{Z})$ . The problem with this transformation is that it does not induce a toral automorphism on  $\mathbb{T}^n$ , hence not a symplectic transformation on  $T^*\mathbb{T}^n$ . To fix this issue, we lift  $T^*\mathbb{T}^n$  to its universal cover  $\mathbb{R}^{2n}$  and interpret this symplectic transformation as that on  $\mathbb{R}^{2n}$ . We will introduce an undoing-the-shear transformation later, so that the composition of the transformations is on  $T^*\mathbb{T}^n$ .

**Lemma 3.1.** *There is a linear symplectic transformation denoted by  $\mathfrak{S}\mathfrak{M}''\mathfrak{M}'$  such that the Hamiltonian system  $\mathbf{H} \circ \phi$  in Proposition 2.2 is reduced to the following  $C^{r-2}$ -Hamiltonian defined on  $\mathbb{R}^{2n}$ , up to an additive constant*

$$(3.4) \quad \begin{aligned} \mathbf{H}_S &:= \mathfrak{S}^{-1*} \mathfrak{M}''^{-1*} \mathfrak{M}'^{-1*} \mathbf{H} \circ \phi \\ &= \tilde{\mathbf{G}}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) + \hat{\mathbf{G}}(\hat{\mathbf{y}}_{n-2}) + P_I(\mathbf{y}) + \delta R_I(\mathbf{x}) + \varepsilon^\sigma R_{II}(\mathbf{x}, \mathbf{y}), \end{aligned}$$

where

$$(3.5) \quad \begin{aligned} \tilde{\mathbf{G}}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) &= \frac{1}{2} \langle \tilde{A} \tilde{\mathbf{y}}, \tilde{\mathbf{y}} \rangle + V(\tilde{\mathbf{x}}) \\ \hat{\mathbf{G}}(\hat{\mathbf{y}}_{n-2}) &= \frac{1}{2} \langle \hat{\mathbf{y}}_{n-2}, B \hat{\mathbf{y}}_{n-2} \rangle + \frac{1}{\sqrt{\varepsilon}} \langle \hat{\omega}_{S, n-2}, \hat{\mathbf{y}}_{n-2} \rangle \end{aligned}$$

where  $\omega_S = (\tilde{0}, \hat{\omega}_{n-2}) = SM''M'\omega_{a''} = M''M'\omega_{a''}$ , for  $S$  in (3.12). The two matrices  $\tilde{A}$  and  $B = (\hat{A} - \check{A}^t \tilde{A}^{-1} \check{A})$  are positive where  $\tilde{A}, \check{A}, \hat{A}$  in  $\mathbb{R}^{2^2}, \mathbb{R}^{2 \times (n-2)}, \mathbb{R}^{(n-2)^2}$  respectively form the matrix

$$(3.6) \quad A = \begin{pmatrix} \tilde{A} & \check{A} \\ \check{A}^t & \hat{A} \end{pmatrix}.$$

The remainders  $R_i(\mathbf{y}) = \mathfrak{S}^{-1*} \mathfrak{M}''^{-1*} \mathfrak{M}'^{-1*} R_i$ ,  $i = I, II, P_I, R_I, R_{II}$  are in (2.19), with  $|R_I(\mathbf{x})|_{C^{r-2}}, |R_{II}(\mathbf{x})|_{C^{r-2}} = O(1)$  as  $\varepsilon, \delta \rightarrow 0$ .

*Proof.* We use the linear symplectic transformation denoted by  $\mathfrak{M}''\mathfrak{M}'$ ,

$$(3.7) \quad x'' = M''M'x, \quad Y'' = M''^{-t}M'^{-t}Y.$$

We also keep track of the frequency vector  $\omega_a'' = M''M'\omega_a = M''\omega_a' = (\nu(a), 0, *, \dots, *)$  where  $\nu(a)$  depends on  $a$  linearly and  $\nu(a'') = 0$ .

We get a Hamiltonian system

$$(3.8) \quad \begin{aligned} H_\delta'' &:= (\mathfrak{M}''^{-1})^* (\mathfrak{M}'^{-1})^* \mathbf{H} \circ \phi \\ &= \frac{1}{\sqrt{\varepsilon}} \langle \omega''^*, Y'' \rangle + \frac{1}{2} \langle A'' Y'', Y'' \rangle + V(x_1'', x_2'') + P_I''(y'') \\ &\quad + \frac{1}{\varepsilon} h(Y^*) + \delta R_I''(x'') + \varepsilon^\sigma R_{II}''(x'', Y'') \end{aligned}$$

by applying  $\mathfrak{M}''\mathfrak{M}'$  term by term to (2.19). In the following discussion, for the simplicity of notations and without leading to confusion, we also remove the  $''$ . Let us denote

$$(3.9) \quad \mathbf{G}(Y, x) = \frac{1}{\sqrt{\varepsilon}} \langle \omega(y^*), Y \rangle + \frac{1}{2} \langle AY, Y \rangle + V(x_1, x_2),$$

where the first two entries of  $\omega(y^*)$  are equal to zero since we have fixed the base point of homogenization  $Y^* \in \Sigma''$  before Proposition 2.2. In this system, the variables  $\hat{Y}_{n-2}$  are cyclic so we can treat this system as a product of an integrable system with a system of two degrees of freedom. Moreover, we can separate variables by introducing a shear transformation.

We write the matrix  $A$  in block form of (3.6). We also denote  $\tilde{v} = (v_1, v_2)$  as the first two entries of a vector  $v \in \mathbb{R}^n$ . Next we have the following formal derivations

$$\begin{aligned}
(3.10) \quad \mathbf{G}(Y, x) &= \frac{1}{\sqrt{\varepsilon}} \langle \omega, Y \rangle + \frac{1}{2} \langle AY, Y \rangle + V(\tilde{x}) \\
&= \frac{1}{2} \langle \tilde{A} \tilde{Y}, \tilde{Y} \rangle + \langle \tilde{Y}, \tilde{A} \hat{Y}_{n-2} \rangle + V(\tilde{x}) \\
&\quad + \frac{1}{2} \langle \hat{A} \hat{Y}_{n-2}, \hat{Y}_{n-2} \rangle + \frac{1}{\sqrt{\varepsilon}} \langle \hat{\omega}_{n-2}, \hat{Y}_{n-2} \rangle \\
&= \frac{1}{2} \langle \tilde{A}(\tilde{Y} + \tilde{A}^{-1} \tilde{A} \hat{Y}_{n-2}), (\tilde{Y} + \tilde{A}^{-1} \tilde{A} \hat{Y}_{n-2}) \rangle + V(\tilde{x}) \\
&\quad - \frac{1}{2} \langle \tilde{A} \hat{Y}_{n-2}, \tilde{A}^{-1} \tilde{A} \hat{Y}_{n-2} \rangle + \frac{1}{2} \langle \hat{A} \hat{Y}_{n-2}, \hat{Y}_{n-2} \rangle + \frac{1}{\sqrt{\varepsilon}} \langle \hat{\omega}_{n-2}, \hat{Y}_{n-2} \rangle \\
&= \frac{1}{2} \langle \tilde{A}(\tilde{Y} + \tilde{A}^{-1} \tilde{A} \hat{Y}_{n-2}), (\tilde{Y} + \tilde{A}^{-1} \tilde{A} \hat{Y}_{n-2}) \rangle + V(\tilde{x}) \\
&\quad + \frac{1}{2} \langle \hat{Y}_{n-2}, (\hat{A} - \tilde{A}^t \tilde{A}^{-1} \tilde{A}) \hat{Y}_{n-2} \rangle + \varepsilon^{-1/2} \langle \hat{\omega}_{n-2}, \hat{Y}_{n-2} \rangle.
\end{aligned}$$

We perform the following linear shear symplectic transformation denoted by  $\mathfrak{S}$ ,

$$(3.11) \quad \begin{bmatrix} \tilde{y} \\ \hat{y}_{n-2} \end{bmatrix} = \begin{bmatrix} \text{id}_2 & \tilde{A}^{-1} \tilde{A} \\ 0 & \text{id}_{n-2} \end{bmatrix} \begin{bmatrix} \tilde{Y} \\ \hat{Y}_{n-2} \end{bmatrix}, \\
\begin{bmatrix} \tilde{x} \\ \hat{x}_{n-2} \end{bmatrix} = \begin{bmatrix} \text{id}_2 & 0 \\ -\tilde{A}^t \tilde{A}^{-t} & \text{id}_{n-2} \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \hat{x}_{n-2} \end{bmatrix}$$

so that the homogenized system in the new coordinates is written in the form  $\mathbf{G} = \tilde{\mathbf{G}} + \hat{\mathbf{G}}$  stated in the lemma. We denote

$$(3.12) \quad S = \begin{bmatrix} \text{id}_2 & 0 \\ -\tilde{A}^t \tilde{A}^{-t} & \text{id}_{n-2} \end{bmatrix}, \quad S^{-t} = \begin{bmatrix} \text{id}_2 & \tilde{A}^{-1} \tilde{A} \\ 0 & \text{id}_{n-2} \end{bmatrix}$$

so that the above symplectic transformation  $\mathfrak{S}$  simplifies to  $\mathbf{x} = Sx$ ,  $\mathbf{y} = S^{-t}Y$ .

Since  $A$  is positive definite and the linear symplectic transformation  $\mathfrak{S}$  does not change the signature so we get both  $\tilde{A}$  and  $B = (\hat{A} - \tilde{A}^t \tilde{A}^{-1} \tilde{A})$  are positive definite.  $\square$

**Remark 3.1.** Notice the above matrix  $S$  is identity in the  $\tilde{x}$  component, hence the Hamiltonian  $\tilde{\mathbf{G}}$  depends on  $\tilde{x}$  periodically. We can project the domain of  $\tilde{\mathbf{G}}$  from  $\mathbb{R}^4$  to  $T^*\mathbb{T}^2$ . So in the following, we think  $\tilde{\mathbf{G}}$  as a Hamiltonian defined on  $T^*\mathbb{T}^2$ .

**3.4. NHIC in Hamiltonian systems of two degrees of freedom.** The system (3.4) is defined in a  $O(\varepsilon^{\sigma-1/2})$ -neighborhood of the sub-manifold of double resonance. We pick a large number  $\Lambda$  independent of  $\varepsilon$  and consider two different regimes in the space of action variables: the low energy region and the high energy region. The low energy region is defined to be a  $\Lambda\varepsilon^{1/2}$ -neighborhood of the submanifold  $\Sigma''$  (2.8) in the original scale (a  $\Lambda$ -neighborhood in the homogenized system). The complement, i.e. the part of action space that is  $\Lambda\varepsilon^{1/2}$  away and within  $\varepsilon^\sigma$  distance from  $\Sigma''$ , is called the high energy region.

In the low energy region, we consider a covering in the following way. In the system  $H_S$  of (3.4), we cover the low energy region by balls of radius  $2\Lambda$  centered at some point  $y^* \in \Sigma''$ . we perform the Taylor expansion around  $y^*$ . Formally we get the same expression as before with slight modification of the values of  $A$  and  $\hat{\omega}_{n-2}$  by  $O(\varepsilon^\sigma)$ , however, the new  $P_1(y)$  term, i.e. the third order and higher terms in  $y - y^*$ , now has estimate  $|P_I|_{C^r} \leq \sqrt{\varepsilon}$  since  $|y - y^*| < \Lambda$  and there is a  $\sqrt{\varepsilon}$  factor in  $P_I$ . So we absorb this  $P_I(y)$  term into the  $\varepsilon^\sigma R_{II}$  terms in (2.19). Later we will allow  $\Lambda$  to depend on  $\delta$ , in which case, the estimate  $|P_I|_{C^r} \leq \sqrt{\varepsilon}$  still holds as  $\varepsilon \rightarrow 0$ . So in the following discussion for the low energy region, we forget about the  $P_I$  term.

We cite the following theorem from [C12] (see also [C15b])

**Theorem 3.3** (Theorem 3.1 of [C12]). *Consider a  $C^2$  Tonelli Hamiltonian  $H$  defined on  $T^*\mathbb{T}^n$ . Given a class  $c_0 \in H^1(\mathbb{T}^n, \mathbb{R})$ , if the minimal measure is supported on a hyperbolic fixed point, then there exists an  $n$ -dimensional flat  $\mathbb{F}_0 \subset H^1(\mathbb{T}^n, \mathbb{R})$  such that this point supports a  $c$ -minimal measure for all  $c \in \mathbb{F}_0$ .*

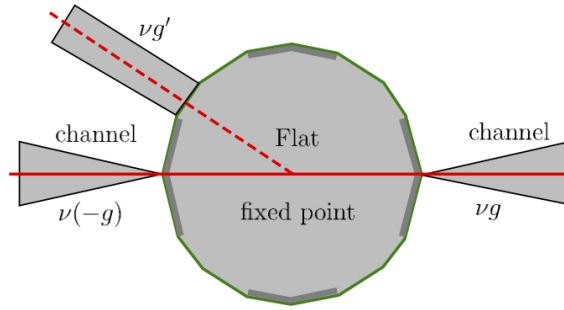


FIGURE 3. Two ways that the flat  $\mathbb{F}_0$  connects to the channels

In the following, we specialize to the case of  $n = 2$ . Each critical point  $\tilde{x}$  of the function  $V$  corresponds to a fixed point  $(x, 0)$  of the Hamiltonian flow  $\Phi_{\tilde{G}}^t$ . Especially, each non-degenerate maximal point determines a hyperbolic fixed point. Let  $J$  denote the standard symplectic matrix, it is obviously a generic property that

**(H2.1):** *The function  $V$  attains its maximum  $\max V = 0$  at one point  $\tilde{x} = 0$  only, where the Hessian matrix of  $V$  is negative definite. All eigenvalues of the matrix  $J \text{diag}(\tilde{A}, D^2V)$  are different:  $-\lambda_2 < -\lambda_1 < 0 < \lambda_1 < \lambda_2$ .*

Under this condition, for fixed  $\hat{y}_{n-2}$ , the system  $\tilde{G}$  is a Hamiltonian system of two degrees of freedom. The  $\alpha$ -function of  $\tilde{G}$  has a 2-dimensional flat  $\mathbb{F}_0$  as the set of its minimal points. For each class  $\tilde{c} \in \text{int}\mathbb{F}_0$ , the minimal measure is uniquely supported on the fixed point, which has at least three minimal homoclinic orbits with different homological class. These homoclinic orbits lie on the intersection of the stable and unstable manifold of the fixed point. As the Hamiltonian  $\tilde{G}$  is autonomous, the intersection is not transversal. However, for convenience, we still call the intersection *transversal* if the tangent space of the stable manifold and that of the unstable manifold span the tangent space of the energy level at the intersection point.

We denote by  $\Lambda_i^+ = (\Lambda_{xi}, \Lambda_{yi})$  the eigenvector corresponding to the eigenvalue  $\lambda_i$ , where  $\Lambda_{xi}$  and  $\Lambda_{yi}$  are for the  $\tilde{x}$ - and  $\tilde{y}$ -coordinate respectively, then the eigenvector for  $-\lambda_i$  will be  $\Lambda_i^- = (\Lambda_{xi}, -\Lambda_{yi})$ . It is also a generic condition that

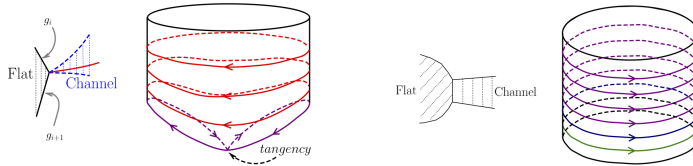
**(H2.2):** *The stable manifold of the fixed point  $(0, 0)$  intersects the unstable manifold transversally along each minimal homoclinic orbit. Each minimal homoclinic orbit approaches to the fixed point along the direction  $\Lambda_1: \dot{\gamma}(t)/\|\dot{\gamma}(t)\| \rightarrow \Lambda_{x1}$  as  $t \rightarrow \pm\infty$ .*

Under these two conditions **(H2.1, H2.2)**, certain uniform hyperbolicity on minimal periodic orbits is proved in Lemma 4.1 (for energy close to zero) and Appendix B of [C12] and [CZ1, C15a] (for all energy levels).

**Theorem 3.4** (Lemma 4.1 and Theorem B.2 of [C12]). *Consider a  $C^r$  ( $r \geq 5$ ) generic Tonelli Hamiltonian  $H$  with two degrees of freedom. Suppose on the zero energy level set, there is a hyperbolic fixed point supporting an action minimizing invariant measure and assume **(H2.1, H2.2)** for the fixed point. Given a class  $g \in H_1(\mathbb{T}^2, \mathbb{Z})$  and any  $\nu^* > 0$ , there are finitely many  $\nu_i \in (0, \nu^*]$  only such that for  $\nu = \nu_i$  there exist exactly two  $\nu g$ -minimal periodic orbits, for other  $\nu \in (0, \nu^*]$  there is only one  $\nu g$ -minimal periodic orbit. These periodic orbits are simultaneously hyperbolic, they make up one or more pieces of normally hyperbolic invariant cylinder.*

In particular, we obtain the generic property that, for the class  $g = (1, 0)$ , all  $\nu g$ -minimal periodic orbits are hyperbolic simultaneously for all  $\nu \in [\nu_0, \nu_1]$  ( $0 < \nu_0 < \nu_1 < +\infty$ ). This Hamiltonian  $\tilde{G}$  as well as the Lagrangian  $\tilde{L}$  (obtained from  $\tilde{G}$  by the Legendre transformation) has been extensively studied in our previous work [C12]. Let us briefly review some results in [C12] if the potential  $V$  satisfies generic conditions **(H2.1, H2.2)**.

- (1) each first homology class  $g \in H_1(\mathbb{T}^2, \mathbb{Z})$  determines a channel in  $H^1(\mathbb{T}^2, \mathbb{R})$ ,  $\mathbb{C}(g) = \mathcal{L}_\beta(\nu g)$  where  $\mathcal{L}_\beta$  is the Legendre-Fenchel transform, connected to the flat  $\mathbb{F}_0$  such that for each  $\tilde{c} \in \mathbb{C}(g)$  the Mather set  $\tilde{\mathcal{M}}(\tilde{c})$  consists of periodic orbits  $\{\tilde{\gamma}, \tilde{\gamma}\}$  with  $[\tilde{\gamma}] = g$ . This channel admits a foliation of lines  $\mathbb{C}(g) = \cup I_E$  with  $I_E \subset \alpha_{\tilde{L}}^{-1}(E)$  such that  $\mathcal{M}(\tilde{c}) = \mathcal{M}(\tilde{c}')$  if  $\tilde{c}, \tilde{c}' \in I_E$  for some  $E \in [0, \infty)$ . For any  $E^* > 0$ , there are finitely many  $E_i \in (0, E^*)$  such that the Mather set consists of two periodic orbits if the first cohomology class  $\tilde{c} \in I_{E_i}$  and the Mather set contains exactly one periodic orbit  $(\tilde{\gamma}_E, \dot{\tilde{\gamma}}_E)$  if the cohomology class does not lie on these lines  $\{I_{E_i}\}$ . All these periodic orbits are hyperbolic; (See **(H4)** and Appendix B of [C12].)
- (2) there are two ways that the channel connected to the flat, either at a point  $\tilde{c}(g)$  or along an edge  $\mathbb{E}(g)$ . In former case, the Mañé set  $\tilde{\mathcal{N}}(\tilde{c}(g))$  consists of the fixed point and two homoclinic orbits  $(\tilde{\gamma}_i, \dot{\tilde{\gamma}}_i)$  and  $(\tilde{\gamma}_j, \dot{\tilde{\gamma}}_j)$ , the periodic orbit  $(\tilde{\gamma}_E, \dot{\tilde{\gamma}}_E)$  approaches these two homoclinic orbits as  $E \rightarrow 0$ . In this case, two integers  $k_i, k_j \in \mathbb{N} \setminus \{0\}$  exists such that  $g = k_i[\tilde{\gamma}_i] + k_j[\tilde{\gamma}_j]$ . In latter case, the periodic orbit  $(\tilde{\gamma}_E, \dot{\tilde{\gamma}}_E)$  approaches a homoclinic orbit or the period remains bounded as  $E \rightarrow 0$ . (See Section 3.3 of [C12].)



**3.5. NHIC in the intermediate region between single and strong double resonance.** In this section, we deal with the high energy region. We need to take the



term  $P_I$  in (2.19) into consideration. Consider the part in (3.4)

$$(3.13) \quad \tilde{G} + P_I(y) = \frac{1}{2} \langle \tilde{A}\tilde{y}, \tilde{y} \rangle + V(\tilde{x}) + P_I(y).$$

We fix  $\hat{y}_{n-2}$  as parameters and apply Theorem 3.4 to  $\tilde{G}$  with the frequency line  $\frac{\partial \tilde{G}}{\partial \tilde{y}} = \nu(1, 0)$ ,  $|\nu| \in [d, \varepsilon^{\sigma-1/2}]$  for some large  $d > 0$  independent of  $\varepsilon$ . Namely  $\tilde{y} \in \Gamma_\nu := (\frac{\partial}{\partial \tilde{y}}(\tilde{G} + P_I))^{-1}(\nu(1, 0))$ . We cover the curve  $\Gamma_\nu$  using balls of radius  $C$  centered on  $\Gamma_\nu$  for some large constant  $C$  independent of  $\varepsilon$ . We denote by  $y^* \in \Gamma_\nu$  any one of such centers and consider the Taylor expansion of  $\tilde{G}$  in a neighborhood of  $y^*$ . Denoting  $y = \tilde{y} - \tilde{y}^*$ ,  $x = \tilde{x}$ , we get a mechanical Hamiltonian of two degrees of freedom up to a term depending on  $\hat{y}_{n-2}$  only,

$$\tilde{G}(x, y) = \frac{\omega_1 y_1}{\sqrt{\varepsilon}} + \frac{1}{2} \langle A' y, y \rangle + V(x) + P'_I(y, \hat{y}_{n-2}), \quad (x, y) \in \mathbb{T}^2 \times \mathbb{R}^2$$

where  $\omega_1 \in [0, \varepsilon^\sigma]$  and  $P'_I$  is the third and higher order terms of the Taylor expansion of  $P_I$  in  $y$ . By the choice of  $y^*$  we have  $|y^* - y^*| \leq \varepsilon^{\sigma-\frac{1}{2}}$ . Clearly, we have

$$A' = \tilde{A} + \frac{\partial^2 P_I}{\partial \tilde{y}^2}(y^*) := \begin{bmatrix} A'_{11} & A'_{12} \\ A'_{12} & A'_{22} \end{bmatrix}.$$

Because of the special form of  $P_I$  (see (2.14) from which  $P_I$  is obtained by a linear transformation in Lemma 3.1), each entry of  $\frac{\partial^2 P_I}{\partial \tilde{y}^2}(y^*)$  is bounded by  $O(\varepsilon^\sigma)$  provided  $|y^* - y^*| \leq \varepsilon^{\sigma-\frac{1}{2}}$ . Moreover, since  $|\hat{y}_{n-2}|$  can be as large as  $\varepsilon^{\sigma-1/2}$ , we also cover the  $n-2$  dimensional space of  $\hat{y}_{n-2}$  using balls of radii  $C$ . Restricted in each such balls, we assume  $|\hat{y}_{n-2}| < C$  in  $P'_I$ , hence  $|P'_I|_{C^r} = O(\sqrt{\varepsilon})$  as  $\varepsilon \rightarrow 0$ .

We first apply Theorem 3.4 to get a normally hyperbolic invariant cylinder. More details are given in [C15a].

We need to make sure such hyperbolicity is uniform for all  $\varepsilon > 0$  so that we can add perturbations by applying the NHIM theorem. The content of the next lemma is explained in the example  $H(x, y) = \frac{y_1^2}{2} + \frac{y_2^2}{2} + (\cos x_2 - 1)$ , for which the cylinder with homology class  $(1, 0)$  is given by  $\{(x_1, y_1, x_2 = 0, y_2 = 0)\}$  whose normal Lyapunov exponents are  $\pm 1$  regardless how high the energy level is. The key idea in the proof of the next lemma is that for high energy level the  $x_1$  variable in  $\tilde{G}$  is fast rotating hence is averaged out and  $y_1$  becomes nearly constant. The remaining system for  $x_2, y_2$  is a system of one degree of freedom which generically admits a hyperbolic fixed point, whose hyperbolicity is independent of  $\varepsilon$ . The normal hyperbolicity is given by the hyperbolic fixed point of the subsystem  $\frac{1}{2}A'_{22}y_2^2 + \int_{\mathbb{T}^1} V(x_1, x_2) dx_1$ .

**Lemma 3.2.** *Consider (3.13). For all parameters  $\hat{y}_{n-2}$  with  $|\hat{y}_{n-2}| \leq \varepsilon^{\sigma-1/2}$ , for  $C^5$  generic  $V$ , the Lyapunov exponents in the normal direction of NHIC given by Theorem 3.4 is bounded uniformly away from zero for all  $|\nu| \in [D, \varepsilon^{\sigma-1/2}]$  for any given large constant  $d$  and for all  $\varepsilon$  sufficiently small. For large enough  $D$ , the NHIC is a whole piece without bifurcation.*

*Proof.* As usual, let  $\Phi_{\tilde{G}}^t$  be the Hamiltonian flow produced by  $\tilde{G}$  and let  $\Phi_{\tilde{G}} = \Phi_{\tilde{G}}^t|_{t=1}$ . What we need to show the cylinder  $\Pi$  that we obtained by applying Theorem 3.4 satisfies Definition 3.1. Since the cylinder is foliated by periodic orbits, the center Lyapunov exponent is zero. We only need to show that the normal Lyapunov exponent is bounded away from zero, independent of  $\varepsilon$ .

Let  $\Omega = \sqrt{\varepsilon}^{-1}\omega_1$ , we consider the case when  $\Omega \gg 1$ . In this case, the energy of the Hamiltonian  $\tilde{G}$  is of order  $O(\Omega)$  if  $\|y\| \leq C$ , where  $C = O(1)$  is independent of  $\varepsilon$ . Under the symplectic coordinate transformation

$$(3.14) \quad (x_1, x_2, y_1, y_2) \rightarrow \left( \Omega x_1, x_2, \frac{y_1}{\Omega}, y_2 \right),$$

the Hamiltonian  $\tilde{G}$  turns out to be

$$\tilde{G} = y_1 + \frac{1}{2\Omega^2} A_{11} y_1^2 + \frac{1}{\Omega} A_{12} y_1 y_2 + \frac{1}{2} A_{22} y_2^2 + V(\Omega x_1, x_2) + P'_I.$$

Before proceeding, we explain the meaning of the rescaling in the  $x_1, y_1$  components. Our goal is to show the normal hyperbolicity of the time-1 map of the original Hamiltonian system  $\tilde{G}$ . Performing the energetic reduction, we want to treat  $y_1$  as the new Hamiltonian and  $x_1$  as the new time. However, since  $|\dot{x}_1|$  is large in the high energy region, we introduce a rescaling such that the rescaled  $x_1$ , considered as the new time has the same time scale as original time. The rescaling is not essential here.

The equation  $\tilde{G}(x_1, y_1(x_1, x_2, y_2), y_2) = E\Omega$  is solved by the function

$$y_1 = E\Omega - \frac{A'_{22}}{2} y_2^2 - A'_{12} E y_2 - V + \Omega^{-1} R_1.$$

Let  $\tau = -x_1$  as the new “time”, Note  $\frac{dx_2}{d\tau} = 1 + O(\Omega^{-1})$ , the normal hyperbolicity of  $\Phi_{\tilde{G}}$  at the energy level set  $\{\tilde{G} = E\Omega\}$  is equivalent to the normal hyperbolicity of  $\Phi_{y_1}^\tau|_{\tau=O(2\pi)}$ . The Hamiltonian  $y_1$  produces a Lagrangian up to an additive constant

$$L_1 = \frac{1}{2A'_{22}} (\dot{x}_2)^2 - \frac{A'_{12}E}{A'_{22}} \dot{x}_2 - V + \Omega^{-1} L_R,$$

where  $L_R$  is  $C^{r-1}$ -bounded in  $(y_2, \tau, x_2)$  for all  $\Omega > 0$ . The minimal periodic orbit of the type- $(\nu, 0)$  for  $\Phi_{\tilde{G}}^t$  is converted to be minimal periodic orbit of  $\phi_{L_1}^\tau$ . As it was shown in [CZ1], the hyperbolicity of such minimal periodic orbit is uniquely determined by the nondegeneracy of the minimal point of following function

$$F(x_2, \Omega, E) = \inf_{\gamma(0)=\gamma([\Omega]\frac{2\pi}{\Omega})=x_2} \int_0^{[\Omega]\frac{2\pi}{\Omega}} L_1(\dot{\gamma}(\tau), \gamma(\tau), \Omega\tau, E) d\tau.$$

As we consider periodic orbits only, the term  $\frac{A'_{12}E}{A'_{22}} \dot{x}_2$  does not contribute to  $F$ , so we drop this term. Let  $\gamma_{\Omega, E}(\tau, x_2)$  be the minimizer of  $F(x_2, \Omega, E)$ , i.e., the action along this curve is equal to  $F(x_2, \Omega, E)$ ,  $|\dot{\gamma}_{\Omega, E}(\tau, x_2)|$  is uniformly bounded for  $\Omega \in [\Omega_0, \infty)$ . As the system has one degree of freedom,  $\Omega^{-1}$ -periodical in  $\tau$ , the minimal point of  $F$  determines an  $\Omega^{-1}$ -periodic curve  $\gamma_{\Omega, E}^*$ . We shall see later that  $|\dot{\gamma}_{\Omega, E}^*(\tau)| \rightarrow 0$  as  $\Omega \rightarrow \infty$ .

Although the Lagrangian  $L_1$  depends on  $\Omega$  in a singular way as  $\Omega \rightarrow \infty$ , the function  $F$  appears regular in  $\Omega$  as  $\Omega \rightarrow 0$ . To see it, let us expand  $V$  into Fourier series

$$V(\Omega\tau, x_2) = -V_0(x_2) + \sum_{k \in \mathbb{Z} \setminus \{0\}} V_k(x_2) e^{ik\Omega\tau}.$$

By the method to prove Riemann-Lebesgue's lemma and using the boundary condition  $\gamma_{\Omega,E}(0, x_2) = \gamma_{\Omega,E}([\Omega] \frac{2\pi}{\Omega}, x_2)$ , we find that for  $k \neq 0$

$$\begin{aligned} \left| \int_0^{[\Omega] \frac{2\pi}{\Omega}} V_k(\gamma_{\Omega,E}(\tau, x_2)) e^{ik\Omega\tau} d\tau \right| &= \left| \frac{1}{ik\Omega} \int_0^{[\Omega] \frac{2\pi}{\Omega}} \dot{V}_k(\gamma_{\Omega,E}(\tau, x_2)) \dot{\gamma}_{\Omega,E}(\tau, x_2) e^{ik\Omega\tau} d\tau \right| \\ &\leq \frac{1}{|k|\Omega} \int_0^{[\Omega] \frac{2\pi}{\Omega}} |\dot{V}_k| |\dot{\gamma}_{\Omega,E}| d\tau \leq \frac{C}{|k|^r \Omega} \end{aligned}$$

where the last inequality is obtained by assuming  $V \in C^r$ . So, for  $r > 2$  we find

$$F(x_2, \Omega, E) = F_0(x_2, \Omega, E) + \Omega^{-1} F_R(x_2, \Omega, E),$$

where

$$F_0 = \int_0^{[\Omega] \frac{2\pi}{\Omega}} \left( \frac{1}{2A'_{22}} (\dot{\gamma}_{\Omega,E}(\tau, x_2))^2 + V_0(\gamma_{\Omega,E}(\tau, x_2)) \right) d\tau.$$

By perturbing the potential  $V \rightarrow V + V_\delta(x_2)$  we can assume generically that  $\frac{d^2 V_0}{dx_2^2} > 0$  holds at the minimal point of  $V_0$ . Consequently, we have  $|\dot{\gamma}_{\Omega,E}^*(\tau)| \rightarrow 0$  as  $\Omega \rightarrow \infty$ .

We claim that some  $\mu > 0$  and  $\Omega_0 > 0$  exist such that  $\partial_{x_2}^2 F \geq \mu$  holds at the minimal point of  $F(\cdot, \Omega, E)$  for all  $\Omega \in [\Omega_0, \infty)$ . As a matter of fact, the minimizer  $\gamma_{\Omega,E}(\tau, x_2)$  is uniquely determined by  $x_2$  provided  $x_2$  stays in certain small neighborhood of the minimal point of  $F(\cdot, \Omega, E)$  (see Section 2 of [CZ1]). It follows that the initial velocity  $\dot{\gamma}_{\Omega,E}(0, x_2)$  smoothly depends on  $x_2$ . Direct calculation shows that

$$\begin{aligned} \frac{\partial^2 F_0}{\partial x_2^2} &= \int_0^{[\Omega] \frac{2\pi}{\Omega}} \left( \frac{1}{A'_{22}} \left( \frac{\partial \dot{\gamma}_{\Omega,E}}{\partial x_2} \right)^2 + \frac{d^2}{dx_2^2} V_0(\gamma_{\Omega,E}^*) \left( \frac{\partial \gamma_{\Omega,E}}{\partial x_2} \right)^2 \right) d\tau \\ &\quad + \int_0^{[\Omega] \frac{2\pi}{\Omega}} \left( \frac{1}{A'_{22}} \dot{\gamma}_{\Omega,E} \frac{\partial^2 \dot{\gamma}_{\Omega,E}}{\partial x_2^2} + \frac{d}{dx_2} V_0(\gamma_{\Omega,E}^*) \frac{\partial^2 \gamma_{\Omega,E}}{\partial x_2^2} \right) d\tau, \end{aligned}$$

where the second integral approaches zero as  $\Omega \rightarrow \infty$ . Indeed, we have  $|\dot{\gamma}_{\Omega,E}^*(\tau)| \rightarrow 0$  as  $\Omega \rightarrow 0$  and  $|\gamma_{\Omega,E}^*(\tau) - x_2^*| \rightarrow 0$  where  $x_2^*$  is the minimal point of  $V_0$ . It follows that  $\frac{dV_0}{dx_2}(\gamma_{\Omega,E}^*(\tau)) \rightarrow 0$  as  $\Omega \rightarrow 0$ . To estimate the first integral, we note that the minimizer  $\gamma_{\Omega,E}^*(\tau)$  stays in small neighborhood of the minimal point of  $V_0$ , certain  $d > 0$  exists (independent of  $\Omega$ ) such that  $\frac{d^2}{dx_2^2} V_0(\gamma_{\Omega,E}^*) \geq d$  holds for all  $\tau \in [0, [\Omega] \frac{2\pi}{\Omega}]$ . As  $\frac{\partial \gamma_{\Omega,E}}{\partial x_2}$  is the solution of the variational equation

$$\frac{1}{2A'_{22}} \frac{d^2}{d\tau^2} \left( \frac{\partial \gamma_{\Omega,E}}{\partial x_2} \right) - \frac{\partial^2 V}{\partial x_2} \frac{\partial \gamma_{\Omega,E}}{\partial x_2} + R = 0$$

with the boundary condition  $\frac{\partial \gamma_{\Omega,E}}{\partial x_2}(0) = \frac{\partial \gamma_{\Omega,E}}{\partial x_2}([\Omega] \frac{2\pi}{\Omega}) = 1$ , where

$$R = R_0 \frac{\partial \gamma_{\Omega,E}}{\partial x_2} + R_1 \frac{d}{d\tau} \frac{\partial \gamma_{\Omega,E}}{\partial x_2} + \varepsilon^\ell R_2 \frac{d^2}{d\tau^2} \left( \frac{\partial \gamma_{\Omega,E}}{\partial x_2} \right)$$

$|R_i|_{C^0}$  is uniformly bounded for  $\Omega > 0$ , certain  $T > 0$  exists (independent of  $\Omega$ ) such that  $\frac{\partial \gamma_{\Omega,E}}{\partial x_2}(\tau) > \frac{1}{2}$  for all  $\tau \in [0, T] \cup [[\Omega] \frac{2\pi}{\Omega} - T, [\Omega] \frac{2\pi}{\Omega}]$ . These arguments lead to the conclusion that certain  $\mu > 0$  exists, independent of  $\Omega$  such that

$$\frac{\partial^2}{\partial x_2^2} F_0(\gamma_{\Omega,E}^*(0), \Omega, E) \geq 2\mu$$

holds for all large  $\Omega \gg 1$ . As the function of action  $F$  is a  $O(\Omega^{-1})$ -perturbation of  $F_0$ , for all large  $\Omega \gg 1$ ,  $\partial_{x_2}^2 F \geq \mu$  holds at the minimal point of  $F(\cdot, \Omega, E)$ .

Given a large  $\Omega$ , let  $x_2^*$  be the minimal point of  $F$  and drop the notation of  $\Omega$ . One has

$$F(x_2, E) - F(x_2^*, E) \geq \mu(x_2 - x_2^*)^2,$$

if  $|x_2 - x_2^*|$  is suitably small. Let  $B_E := u^- - u^+$  denote the barrier function where  $u^\pm$  are the backward and forward weak KAM solutions (see the appendix A), as it was shown in [CZ1], one has

$$B_E(x_2) - B_E(x_2^*) \geq F(x_2, E) - F(x_2^*, E).$$

As barrier function is semi-concave, there exists a number  $C_L > 2\mu$  such that

$$B_E(x_2) - B_E(x_2^*) \leq C_L(x_2 - x_2^*)^2.$$

It guarantees that the hyperbolicity of the minimizer is not weaker than  $\Lambda = \sqrt{1 - \frac{2\mu}{C_L}}$ . Let us assume the contrary, denote by  $(\gamma_E^*(\tau), \dot{\gamma}_E^*(\tau))$  the minimal periodic orbit and denote by  $(\gamma^\pm(\tau), \dot{\gamma}^\pm(\tau))$  the orbit such that  $\gamma^-(0) = \gamma^+(0)$  and they asymptotically approaches to the orbit  $(\gamma^*(\tau), \dot{\gamma}^*(\tau))$  as  $\tau \rightarrow \pm\infty$ , we then have

$$|\gamma_E^*(\pm j) - \gamma^\pm(\pm j)| > \Lambda |\gamma_E^*(\pm(j-1)) - \gamma^\pm(\pm(j-1))|$$

if  $|\gamma_E^*(0) - \gamma^\pm(0)|$  is suitably small. The following computation leads to a contradiction:

$$\begin{aligned} C_L(\gamma^\pm(0) - \gamma_E^*(0))^2 &\geq B_E(\gamma^\pm(0)) - B_E(\gamma_E^*(0)) \\ &\geq \sum_{j=1}^{\infty} \left( F(\gamma^-(j)) - F(\gamma_E^*(0)) \right) + \left( F(\gamma^+(j)) - F(\gamma_E^*(0)) \right) \\ &> 2\mu \frac{(\gamma^\pm(0) - \gamma_E^*(0))^2}{1 - \Lambda^2} = C_L(\gamma^\pm(0) - \gamma_E^*(0))^2. \end{aligned}$$

Since  $E$  is bounded uniformly for  $\Omega > 0$  and appears in the remainder  $\Omega^{-1}L_R$ , such normal hyperbolicity holds for all  $E$  we are concerned about.

Notice that in the above argument, the normal hyperbolicity for high enough energy level is completely determined by the hyperbolic fixed point of the subsystem  $H_2(x_2, y_2) = \frac{A'_{22}y_2^2}{2} - V_0(x_2)$  where  $V_0(x_2) = -\int_{\mathbb{T}^1} V(x_1, x_2) dx_1$  and  $|A_{22} - A'_{22}| = O(\varepsilon^\sigma)$  for all  $|\hat{y}_{n-2}| < \varepsilon^{\sigma-1/2}$ . We get the uniform normal hyperbolicity for all  $|\hat{y}_{n-2}| < \varepsilon^{\sigma-1/2}$ . For generic  $V$ , the function  $V_0$  has unique non degenerate global min, in which case the NHIC stays in a neighborhood of the set  $\{(x_1, y_2, x_2 = \argmin V_0, y_2 = 0)\}$  for high enough energy levels, hence the NHIC has no bifurcations.  $\square$

Finally, we formulate the following lemma summarizing the way we apply the NHIM Theorem 3.1 and 3.2.

**Lemma 3.3.** *Suppose the Hamiltonian system  $H$  satisfies **(H1.1)**, **(H1.2)**, **(H2)** along the frequency line segment  $\omega_a = \lambda_a \left( a, \frac{p}{q}, \frac{p}{Q}, \hat{\omega}_{n-3}^* \right)$ . Then there exists  $\Delta(\mathbf{k}') > 0$  depending only on  $\mathbf{k}'$  but independent of  $\delta, \varepsilon$ , being the maximal allowable  $C^1$  norm of the perturbation, such that for each  $\delta < \Delta(\mathbf{k}')$ , where  $\delta$  is in Proposition 2.1 and 2.2, Theorem 3.1 and 3.2 are applicable to the truncated system in a  $\mu$ -neighborhood of the frequency line  $\omega_a$  to get the existence of NHIC in the full system in all the single, weak double resonant, and the strong double resonant case.*

*Proof.* The existence of such  $\Delta(\mathbf{k}')$  in the single resonance and weak double resonance case is given by **(H1.1)**. In the double resonance case, there are sub cases. In the low energy region for  $E < \Lambda$  for some large  $\Lambda$ , the uniform normal hyperbolicity follows

from Theorem 3.4 immediately. In the high energy region  $E > \Lambda$  the uniform normal hyperbolicity is given by Lemma 3.2.  $\square$

**3.6. Center straightening and reduction of Hamiltonian systems with two degrees of freedom.** We consider a Tonelli Hamiltonian system of two degrees of freedom  $H(x, y)$ ,  $(x, y) \in T^*\mathbb{T}^2$  and assume  $(0, 0, 0, 0)$  is a hyperbolic fixed point on the zero energy level supporting the Mather set of rotation vector 0. We fix a class  $g \in H_1(\mathbb{T}^2, \mathbb{Z})$  then consider  $g_\nu := \nu g$  ( $\nu \neq 0$ ). We get that the Mather set  $\mathcal{M}_{g_\nu}$  is supported on a periodic orbits except that for finitely many  $\nu$ 's, the corresponding Mather set is supported on two hyperbolic periodic orbits. We get a cylinder foliated by hyperbolic periodic orbits  $(\dot{\gamma}_\nu, \gamma_\nu)$  as the Mather set of  $\mathcal{M}_{g_\nu}$ . Let  $\Pi_0 = \mathcal{L}_{\tilde{G}}(\cup_\nu \mathcal{M}_{g_\nu})$  (recall  $\mathcal{L}_{\tilde{G}}$  is the Legendre-Fenchel transformation  $TM \rightarrow T^*M$ ). Since each periodic orbit is hyperbolic, we get  $\Pi_0$  is a normally hyperbolic invariant cylinder of dimension two.

The normal hyperbolicity gives rise to the following decomposition of the symplectic form (Equation (63) of [DLS08])

$$(3.15) \quad \Omega = \left[ \begin{array}{c|c|c} 0 & \Omega^{su} & 0 \\ \hline -\Omega^{su} & 0 & 0 \\ \hline 0 & 0 & \Omega|_{E^c} \end{array} \right],$$

with respect to the decomposition of the tangent space

$$T_x M = E_x^s \oplus E_x^u \oplus E_x^c, \quad x \in \Pi_0.$$

Especially, the symplectic form  $\Omega$  restricted to the cylinder is still a symplectic form. Finally, according to Weinstein's neighborhood theorem (Theorem 3.30 of [MS]), we can always find symplectic coordinates  $(I, \varphi, u, v)$  in a neighbourhood of the periodic orbits, such that  $(I, \varphi, 0, 0)$  corresponds to the NHIC. A neighborhood of the NHIM is realized as a normal bundle. The symplectic coordinates  $(I, \varphi)$  in the NHIM is given by the next proposition. After that we consider the fiber coordinates. Notice  $\dim E^u = \dim E^s = 1$ , we can write  $\Omega^{su} = du \wedge dv$  where  $u$  is the unit covector of  $E^u$  and  $v$  is the unit covector of  $E^s$ . We choose  $u, v$  as the fiber coordinates.

The next proposition holds true for all homology classes  $g \in H_1(\mathbb{T}^2, \mathbb{Z})$ . However, when applying the proposition in this paper, we choose exclusively  $g = (1, 0)$ .

**Proposition 3.1.** *For a Hamiltonian  $H \in C^r(T^*\mathbb{T}^2, \mathbb{R})$  and a homological class  $g \in H_1(\mathbb{T}^2, \mathbb{Z})$ , we assume that*

- (1)  $H(0, 0) = 0$ . The Hamiltonian flow  $\Phi_H^t$  takes  $(x, y) = (0, 0)$  as its hyperbolic fixed point, the point  $\mathcal{L}_H(0, 0)$  supports  $c$ -minimal measure for all  $c \in \mathbb{F}_0$ ;
- (2) Some  $E_1 > 0$  exists such that for each  $c \in I_E \subset \mathbb{C}(g)$  with  $E \in (0, E_1)$  the Mather set  $\tilde{\mathcal{M}}(c)$  consists of a unique periodic orbit  $(\dot{\gamma}_E, \gamma_E)$  with  $[\gamma_E] = g$ .

Let  $\Pi_0 = \mathcal{L}_H^{-1} \cup \{(\gamma_E, \dot{\gamma}_E) : E \in (0, E_1)\}$ , then restricted on the cylinder  $\Pi_0$ , there exists a symplectic change of variables  $\Phi : (I, \varphi) \in \mathbb{R} \times \mathbb{T} \rightarrow (x, y)|_{\Pi_0}$ , such that the Hamiltonian  $H$  can be written as

$$\Phi^* H = H \circ \Phi = \tilde{h}(I),$$

where  $\tilde{h} \in C^r$  has the following properties when  $I > 0$

$$\tilde{h}(0) = 0, \quad \frac{d\tilde{h}}{dI}(0) = 0, \quad \frac{d\tilde{h}}{dI} > 0, \quad \frac{d^2\tilde{h}(I)}{dI^2} > 0.$$

Assume furthermore that  $H$  is reversible, i.e.  $H(x, y) = H(x, -y)$ , then  $\tilde{h}(I)$  is also reversible,  $\tilde{h}(I) = \tilde{h}(-I)$ .

*Proof.* Clearly,  $\Pi_0$  is diffeomorphic to a piece of cylinder, which is a symplectic submanifold as one has

$$\int_{\Pi_0} \Omega = \int_0^{E_1} \int_0^{T(E)} dE \wedge dt > 0$$

where  $T(E)$  is the period of the periodic orbit lying on the energy level set  $H^{-1}(E)$ . Let  $\Omega_{\Pi_0}$  be the restriction of standard symplectic form  $\Omega$  on the cylinder. Denoted by  $\Pi = \mathbb{T} \times \mathbb{R}$  the standard cylinder and let  $\Psi_0: \Pi \rightarrow \Pi_0$  be a diffeomorphic map. Then the pull back of  $\Omega_{\Pi_0}$ ,  $\Psi_0^* \Omega_{\Pi_0}$  is a symplectic form on the standard cylinder  $\Pi_0$ . As the second de Rham cohomology group of cylinder  $\Pi$  is trivial, Moser's argument on the isotopy of symplectic forms shows that certain diffeomorphism  $\Psi: \Pi \rightarrow \Pi$  exists such that

$$\Psi^* \Psi_0^* \Omega_{\Pi} = dx \wedge dy.$$

The Hamiltonian  $H$  induces a Hamiltonian defined on  $\Pi$ :  $H\Psi_0\Psi(x, y)$ .

Let us consider a Tonelli Hamiltonian  $H(x, y)$  where  $(x, y) \in \mathbb{T} \times \mathbb{R}$ . By adding a closed 1-form and a constant, we can assume that its  $\alpha$ -function attains its minimum at the zero first cohomology, i.e.  $\alpha_H(0) = \min \alpha_H = 0$ , the Mather set is a singleton of hyperbolic fixed point located at the origin  $(x, y) = (0, 0)$ . For  $E \geq 0$ , the equation

$$(3.16) \quad H(x, W(x, E)) = E, \quad \forall x \in \mathbb{T}$$

has exactly two global solutions  $W^+(x, E) \geq W^-(x, E)$ . These two solutions produce the separatrix  $\Sigma_0^+$  and  $\Sigma_0^-$ :  $\Sigma_0^\pm = \text{graph}(W^\pm(\cdot, 0))$  if  $E = 0$ .

In the region above the separatrix  $\Sigma_0^+$  (or below the separatrix  $\Sigma_0^-$ ), one can introduce new angle-action variable  $(I, \varphi)$  so that  $H(x, y) = E(I)$ . The coordinate transformation  $(x, y) \rightarrow (I, \varphi)$  is determined by the generating function  $S(I, x)$  which solves the Hamilton-Jacobi equation

$$H\left(\frac{\partial S}{\partial x}, x\right) = E(I).$$

Since the Hamiltonian is strictly convex in  $y$ , this equation determines exactly two global solutions if  $E > \min \alpha_H$ , denoted by  $W^+(x, E)$  and  $W^-(x, E)$ . The graph of  $W^+(x, E)$  is above the separatrix  $\Sigma_0^+$ , the graph of  $W^-(x, E)$  below the separatrix  $\Sigma_0^-$ .

Let us consider the region above the separatrix  $\Sigma_0^+$ , where the generating function

$$S(I, x) = \int W^+(x, E(I)) dx,$$

produces a coordinate transformation  $\Psi_1: (x, y) \rightarrow (I, \varphi)$  ( $E$  uniquely determines  $I$ )

$$(3.17) \quad y = \frac{\partial S}{\partial x} = W^+(x, E), \quad \varphi = \frac{\partial S}{\partial I} = \frac{dE}{dI} \int \frac{\partial}{\partial E} W^+(x, E) dx.$$

From Section 50 B and C of [A89], the action variable  $I = I(x, y)$  is defined as the area bounded by the curve  $\text{graph}(W^+(\cdot, E))$  and the curve  $\Sigma_0^+$ ,

$$(3.18) \quad I(E) = \frac{1}{2\pi} \int_0^{2\pi} (W^+(x, E) - W^+(x, 0)) dx.$$

The symplectic change of coordinates is the composition of these maps  $\Phi = \Psi_1 \Psi_0 \Psi$ .

It remains to show the twist. We use a result of Carneiro [Ca] saying that Mather's  $\beta$  function is differentiable in the radial direction for autonomous systems. Now  $\tilde{h}(I)$  is actually Mather's  $\alpha$  function since Mather set is exactly the periodic orbit  $\gamma_\nu$ . The direction of  $\nu g$  is the radial direction as  $\nu$  varies. The  $\alpha$  function is strictly convex  $\frac{d^2 \tilde{h}(I)}{dI^2} > 0$ , a.e. in order that  $\beta$  is differentiable.

$$\frac{d\tilde{h}(I)}{dI} = \frac{d\tilde{h}(0)}{dI} + \int_0^I \frac{d^2 \tilde{h}(t)}{dI^2} dt = \int_0^I \frac{d^2 \tilde{h}(t)}{dI^2} dt > 0.$$

Since the symplectic transformation is explicit, we get that  $\tilde{h} \in C^r$ .

Finally, to see the system  $\tilde{h}(I)$  is reversible, we notice that the reversibility of the system  $H(x, y)$  implies that the Mather sets with rotation vectors  $\nu g$  and  $-\nu g$ ,  $\nu > 0$  are supported on the same periodic orbit with reversed time. Since the Legendre transform of an even function is also even, so we get the Lagrangian  $L(\dot{x}, x)$  is even with respect to  $\dot{x}$ , hence  $p = \frac{\partial L}{\partial \dot{x}}$  get a negative sign we reverse the time. The two periodic orbits lie on the same energy level and their corresponding action variables are opposite to each other from the formula  $I = \frac{1}{2\pi} \oint y dx$ . This completes the proof.  $\square$

**3.6.1. The estimate of the derivatives of the canonical symplectic transformation near strong double resonance.** When we are getting close to the strong double resonance point the symplectic transformation in Proposition 3.1 becomes singular. Let us show how that symplectic transformation in the homogenized coordinates behaves when the energy of the periodic orbit is getting closer and closer to zero.

**Lemma 3.4.** *Consider the Hamiltonian  $H(x, y)$  in Proposition 3.1. Then, for energy level greater than small  $E > 0$ , we have the following estimates for the symplectic transform  $\Phi$  given by Proposition 3.1*

$$(3.19) \quad \left| \frac{\partial(x, y)}{\partial(I, \varphi)} \right| \leq d_1 |\ln E|,$$

$$(3.20) \quad \max_{0 \leq k_1 \leq k} \left\{ \left| \frac{\partial^k x}{\partial \varphi^{k_1} \partial I^{k-k_1}} \right|, \left| \frac{\partial^k y}{\partial \varphi^{k_1} \partial I^{k-k_1}} \right| \right\} \leq \frac{d_k}{E^{k-1+\delta}},$$

where the constant  $\delta > 0$  is small and  $d_k > 0$  is independent of  $E$ .

**Remark 3.2.** Because of this lemma, we consider only the part of phase space with  $|I| > \gamma > 0$  for some  $\gamma$  independent of  $\delta, \varepsilon$ . Hence we have  $E = \tilde{h}(\gamma) > 0$ . This  $\gamma$  will be chosen in Section 6.3.

*Proof.* Recall, in the region above the separatrix  $\Sigma_0^+$ , new action variable  $I = I(x, y)$  is introduced as the area bounded by the curve  $\text{graph}(W^+(\cdot, E))$  and the curve  $\Sigma_0^+$ ,

$$I(E) = \frac{1}{2\pi} \int_0^{2\pi} (W^+(x, E) - W^+(x, 0)) dx.$$

In  $(I, \varphi)$ -coordinates, the speed of  $\varphi$  remains constant and approaches to zero as  $E$  decreases to zero while the speed of  $x$  is bounded away from zero when  $(x, y)$  is not close to the hyperbolic fixed point. This fact reminds us of the singularity of  $E(I)$  as  $E \rightarrow 0$ . Let us explore how singular it would be. Since the function  $W^+$  solves the equation (3.16) for each  $E > 0$ , we have

$$\frac{dI}{dE} = \frac{1}{2\pi} \int_0^{2\pi} \frac{\partial}{\partial E} W^+(x, E) dx = \frac{1}{2\pi} \int_0^{2\pi} \left( \frac{\partial H}{\partial y} \right)^{-1} (x, W^+(x, E)) dx.$$

For each  $E > 0$ ,  $\partial_y H = \dot{x}$  remains away from zero when  $(x, y)$  is not close to the fixed point. For small  $E > 0$ , the main contribution to the quantity  $\frac{dI}{dE}$  is the integral over a small neighborhood of  $x = 0$ . In a small neighborhood of the hyperbolic fixed point the Hamiltonian takes a form

$$H(x, y) = \frac{1}{2}y^2 - \frac{\lambda^2}{2}x^2 + H_3(x, y)$$

where  $H_3(x, y) = O(\|(x, y)\|^3)$ . Therefore, in a small neighborhood of the fixed point, the energy level line  $H^{-1}(E)$  is the graph of the function

$$(3.21) \quad y = \sqrt{2E + \lambda^2 x^2} \left(1 + W_1(x)\right), \quad x \in [-\delta, \delta]$$

where  $W_1(x) = O(|x|)$ . Consequently, we have

$$(3.22) \quad \begin{aligned} \frac{dI}{dE} &= \frac{1}{2\pi} \int_{-\delta}^{\delta} \frac{dx}{\sqrt{2E + \lambda^2 x^2} (1 + W_1(x))} + \frac{1}{2\pi} \int_{\delta}^{2\pi-\delta} \frac{dx}{\partial_y H} \\ &= -\frac{1}{\pi\lambda} \ln E + B_1(E) \end{aligned}$$

where  $B_1$  depends on  $\delta$  and  $\lambda$  but remains bounded as  $E \rightarrow 0$ .

To explore higher order singularity of  $I$  on  $E$ , we do further calculation

$$\begin{aligned} \frac{d^k I}{dE^k} &= \frac{a_k}{2\pi} \int_0^{2\pi} \left( \frac{\partial H}{\partial y} \right)^{-(2k-1)} \left( \frac{\partial^2 H}{\partial y^2} \right)^{k-1} (x, W^+(x, E)) dx \\ &\quad + \frac{1}{2\pi} \int_0^{2\pi} F_{2k-2} \left( \left( \frac{\partial H}{\partial y} \right)^{-1} \right) dx \end{aligned}$$

where  $a_k = (-1)^{k-1} (2k-3)(2k-5) \cdots 1$ ,  $F_{2k-2}$  is a polynomial of  $(2k-2)$ -order in  $\partial_y^{-1} H$ , its coefficients are the functions of the derivatives of  $H$ . To obtain an estimate on the  $k$ -th derivative of  $I$  in  $E$ , we make use of again the formula (3.21) and the fact that  $\partial_y H = \dot{x}$  remains away from zero when  $(x, y)$  is not close to the fixed point. For the case  $k \geq 2$  we find

$$(3.23) \quad \frac{d^k I}{dE^k} = A_k(E) E^{1-k} (1 + B_k(E))$$

where the functions  $A_k$  and  $B_k$  depends on  $\mu$  and  $d$  but remains bounded as  $E \rightarrow 0$ . Indeed, we have

$$A_k(E) = \frac{1}{2^{k-1}\pi} \sum_{\ell=0}^{k-2} \frac{(-1)^\ell}{2\ell+1} \binom{\ell}{k-2} \left( \frac{\mu d}{\sqrt{2E + \mu^2 d^2}} \right)^{2\ell+1}.$$

Guaranteed by the strict monotonicity of  $I$  in  $E$  (see (3.22)), the function  $E = E(I)$  and its inverse  $I = I(E)$  are well defined. The identity  $E(I(E)) = E$  yields

$$\frac{dE}{dI} \frac{dI}{dE} \equiv 1 \Rightarrow \frac{dE}{dI} = \left( \frac{dI}{dE} \right)^{-1},$$

which in turn produces inductively the expression of second derivative

$$\frac{d^2 E}{dI^2} = - \left( \frac{dI}{dE} \right)^{-3} \frac{d^2 I}{dE^2},$$

and higher order of derivatives. We thus obtain that

$$(3.24) \quad \left| \frac{dE}{dI} \right| \leq \frac{d'_1}{|\ln E|}, \quad \left| \frac{d^k E}{dI^k} \right| \leq \frac{d'_k}{|\ln E|^{k+1} E^{k-1}}, \quad \forall k \geq 2,$$



where the constants  $d'_k > 0$  ( $k = 1, 2, \dots$ ) are independent of the energy  $E > 0$ .

With these formulae we are ready to estimate the  $C^k$ -norm of the coordinate transformation  $T: (x, y) \rightarrow (I, \varphi)$ . This transformation is determined by the generating function  $S(I, x)$  which solves the Hamilton-Jacobi equation

$$H\left(\frac{\partial S}{\partial x}, x\right) = E(I).$$

Since the Hamiltonian is strictly convex in  $y$ , this equation determines exactly two global solutions if  $E > \min \alpha_H$ , denoted by  $W^+(x, E)$  and  $W^-(x, E)$ . The graph of  $W^+(x, E)$  is above the separatrix  $\Sigma_0^+$ , the graph of  $W^-(x, E)$  below the separatrix  $\Sigma_0^-$ . These two solutions satisfy an identity

$$(3.25) \quad \frac{\partial H}{\partial y} \frac{\partial W^\pm}{\partial E} = 1.$$

Let us consider the region above the separatrix  $\Sigma_0^+$ , where the generating function

$$S(I, x) = \int W^+(x, E(I)) dx,$$

produces a coordinate transformation  $\Psi_1: (x, y) \rightarrow (I, \varphi)$  ( $E$  uniquely determines  $I$ ) determined by formula (3.17). Restricted in the region where  $E > 0$  we obtain from the identity (3.25) that

$$\frac{\partial^2 S}{\partial I \partial x} = \left(\frac{\partial H}{\partial y}\right)^{-1} \frac{dE}{dI} \neq 0.$$

Therefore, equation (3.17) well defines the coordinate transformation  $(x, y) \rightarrow (I, \varphi)$  such that

$$(3.26) \quad \begin{aligned} \frac{\partial x}{\partial \varphi} &= \left(\frac{dE}{dI}\right)^{-1} \left(\frac{\partial^2 S}{\partial E \partial x}\right)^{-1}, \\ \frac{\partial x}{\partial I} &= - \left( \frac{\partial^2 S}{\partial E^2} \frac{dE}{dI} + \frac{\partial S}{\partial E} \frac{d^2 E}{dI^2} \left(\frac{dE}{dI}\right)^{-1} \right) \left(\frac{\partial^2 S}{\partial E \partial x}\right)^{-1} \\ \frac{\partial y}{\partial \varphi} &= \left(\frac{dE}{dI}\right)^{-1} \frac{\partial^2 S}{\partial x^2} \left(\frac{\partial^2 S}{\partial E \partial x}\right)^{-1}, \\ \frac{\partial y}{\partial I} &= \frac{\partial^2 S}{\partial E \partial x} \left( \frac{dE}{dI} + \frac{\partial^2 S}{\partial x^2} \left(\frac{dE}{dI}\right)^{-1} \right). \end{aligned}$$

Indeed, these equations are derived as follows. As  $I$  uniquely determines  $E$ , we can write the generating function  $S(I, x) = S(E, x)$ . For the equation  $\varphi = \frac{\partial S}{\partial I}$ , think  $x, y$  as the function of  $I, \varphi$  and take differential for  $\varphi$  on both sides, we have

$$1 = \frac{\partial^2 S}{\partial I \partial x} \frac{\partial x}{\partial \varphi} = \frac{\partial^2 S}{\partial E \partial x} \frac{\partial x}{\partial \varphi} \frac{dE}{dI}.$$

From which we obtain

$$\frac{\partial x}{\partial \varphi} = \left(\frac{\partial^2 S}{\partial E \partial x}\right)^{-1} \left(\frac{dE}{dI}\right)^{-1}.$$

Take differential for  $I$  we have

$$0 = \frac{\partial^2 S}{\partial E^2} \left(\frac{dE}{dI}\right)^2 + \frac{\partial S}{\partial E} \frac{d^2 E}{dI^2} + \frac{\partial^2 S}{\partial E \partial x} \frac{dE}{dI} \frac{\partial x}{\partial I}.$$

from which we obtain the second equation. From the equation  $y = \frac{\partial S}{\partial x}$ , we obtain

$$\frac{\partial y}{\partial \varphi} = \frac{\partial^2 S}{\partial x^2} \frac{\partial x}{\partial \varphi} = \frac{\partial^2 S}{\partial x^2} \left( \frac{\partial^2 S}{\partial E \partial x} \right)^{-1} \left( \frac{dE}{dI} \right)^{-1},$$

that is the third equation. By the same method, we can get the fourth equation. With the help of these formulae as well as (3.24) we finally obtain the estimate stated in the lemma.  $\square$

**3.7. Reduction of order and undoing the shear near strong double resonance in the full system.** In this section, we perform the reduction of order near strong double resonance in the full system. There is one special thing that we need to do is to undo the symplectic transformations  $\mathfrak{S}$  since the matrix  $S$  is not necessarily unimodular matrix of integer and may destroy the resonance relations.

Begin with the system (3.5), we apply Proposition 3.1 to  $(x_1, x_2, y_1, y_2)$ -component with homology class  $g = (1, 0)$  to get new coordinates  $k_\delta(I, \varphi, u, v) = (x_1, x_2, y_1, y_2)$  so that the cylinder takes the form  $\{(I, \varphi) \in \mathbb{R} \times \mathbb{T}, (u, v) = 0\}$ . Since  $(u, v) = 0$  on the invariant cylinder, where  $k_\delta$  is the restriction map in Theorem 3.2, we have

$$(3.27) \quad \begin{aligned} \tilde{x}(I, \varphi, u, v) &= \tilde{x}^*(I, \varphi) + \tilde{x}_r(I, \varphi, u, v), \\ \tilde{y}(I, \varphi, u, v) &= \tilde{y}^*(I, \varphi) + \tilde{y}_r(I, \varphi, u, v), \end{aligned}$$

where  $|\tilde{x}_r|/\|(u, v)\| \rightarrow 0$  and  $|\tilde{y}_r|/\|(u, v)\| \rightarrow 0$  as  $\|(u, v)\| \rightarrow 0$ . Extending  $k_\delta$  in natural way to all variables, denoted by  $\mathfrak{K}_\delta(I, \varphi, u, v, \hat{x}_{n-2}, \hat{y}_{n-2}) = (\tilde{x}, \tilde{y}, \hat{x}_{n-2}, \hat{y}_{n-2})$ , we get the Hamiltonian  $\mathfrak{K}_\delta^* H_S$  in the variables  $(I, u, \hat{y}_{n-2}; \varphi, v, \hat{x}_{n-2})$ , where  $H_S$  is in (3.4). We have  $H_S = (\mathfrak{S}^{-1})^* H_\delta''$  from (3.4), where  $H_\delta''$  is in (3.8). The variables  $(x'', Y'')$  (we omit '' for them in the following) of  $H_\delta''$  undergo the following transformations by  $\mathfrak{K}_\delta \mathfrak{S}$  (see (3.12) for  $\mathfrak{S}$ )

$$(3.28) \quad \begin{aligned} \begin{bmatrix} \tilde{x} \\ \hat{x}_{n-2} \end{bmatrix} &= \begin{bmatrix} \text{id}_2 & 0 \\ \tilde{A}^t \tilde{A}^{-t} & \text{id}_{n-2} \end{bmatrix} \begin{bmatrix} \tilde{x}(I, v, \varphi, u) \\ \hat{x}_{n-2} \end{bmatrix}, \\ \begin{bmatrix} \tilde{Y} \\ \hat{Y}_{n-2} \end{bmatrix} &= \begin{bmatrix} \text{id}_2 & -\tilde{A}^{-1} \tilde{A} \\ 0 & \text{id}_{n-2} \end{bmatrix} \begin{bmatrix} \tilde{y}(I, v, \varphi, u) \\ \hat{y}_{n-2} \end{bmatrix}. \end{aligned}$$

Put Formula (3.28) into the fourth equality of (3.10), notice  $\hat{y}_{n-2} = \hat{Y}_{n-2}$  and apply Theorem 3.2 and Proposition 3.1 we get a Hamiltonian

$$(3.29) \quad \begin{aligned} \mathfrak{K}_\delta^* H_S &= \frac{1}{2} \langle \tilde{A} \tilde{y}(I, \varphi, u, v), \tilde{y}(I, \varphi, u, v) \rangle + V(\tilde{x}(I, \varphi, u, v)) \\ &\quad + P_I(\tilde{y}(I, \varphi, u, v), \hat{y}_{n-2}) + \frac{1}{2} \langle \hat{y}_{n-2}, B \hat{y}_{n-2} \rangle + \frac{1}{\sqrt{\varepsilon}} \langle \hat{\omega}_{n-2}, \hat{y}_{n-2} \rangle \\ &\quad + \mathfrak{K}_\delta^* (\delta R_I(x) + \varepsilon^\sigma R_{II}(x, y)) \\ &= \tilde{h}(I) + \frac{1}{\sqrt{\varepsilon}} \langle \hat{\omega}_{n-2}, \hat{y}_{n-2} \rangle + \mathfrak{K}_\delta^* P_I(I, \varphi, u, v, \hat{y}_{n-2}) \\ &\quad + \frac{1}{2} \langle \hat{y}_{n-2}, B \hat{y}_{n-2} \rangle + \mathfrak{K}_\delta^* (\delta R_I(x) + \varepsilon^\sigma R_{II}(x, y)) \\ &\quad + \left[ \frac{1}{2} \langle \tilde{A} \tilde{y}(I, \varphi, u, v), \tilde{y}(I, \varphi, u, v) \rangle - \frac{1}{2} \langle \tilde{A} \tilde{y}^*(I, \varphi), \tilde{y}^*(I, \varphi) \rangle \right] \\ &\quad + [V(\tilde{x}(I, \varphi, u, v)) - V(\tilde{x}^*(I, \varphi))]. \end{aligned}$$

By formula (3.27), the sum of the last four terms in the bracket in above formula, as well as its first derivatives, is equal to zero when  $u = v = 0$ . By the condition assumed

the set

$$\Pi = \{(I, \varphi, u, v, \hat{x}_{n-2}, \hat{y}_{n-2} : (u, v) = 0\}$$

is a NHIC for the Hamiltonian flow of  $\mathfrak{K}_\delta^* \mathbf{H}_S$ . Restricted on the NHIC, the above system is a system of  $n - 1$  degrees of freedom of the following form

$$(3.30) \quad \begin{aligned} k_\delta^* \mathbf{H}_S = & \tilde{h}(I) + \frac{1}{\sqrt{\varepsilon}} \langle \hat{\omega}_{n-2}, \hat{y}_{n-2} \rangle + k_\delta^* P_I(I, \varphi, \hat{y}_{n-2}) \\ & + \frac{1}{2} \langle \hat{y}_{n-2}, B \hat{y}_{n-2} \rangle + k_\delta^* (\delta R_I(x) + \varepsilon^\sigma R_{II}(x, y)) \end{aligned}$$

where  $k_\delta^*$  is the restriction of  $\mathfrak{K}_\delta$  to the NHIC.

Next let us introduce another symplectic coordinate transformation (undoing the transformation  $\mathfrak{S}$ ) since in the current coordinates the system is not necessarily periodic in  $\hat{x}_{n-2}$  due to the shear transformation  $\mathfrak{S}$ ,

$$(3.31) \quad \begin{aligned} \begin{bmatrix} \psi \\ \mathbf{u} \\ \hat{\mathbf{X}}_{n-2} \end{bmatrix} &= \begin{bmatrix} \text{id}_2 & 0 \\ \check{A}^t \check{A}^{-t} & \text{id}_{n-2} \end{bmatrix} \begin{bmatrix} \varphi \\ u \\ \hat{x}_{n-2} \end{bmatrix}, \\ \begin{bmatrix} J \\ \mathbf{v} \\ \hat{\mathbf{Y}}_{n-2} \end{bmatrix} &= \begin{bmatrix} \text{id}_2 & -\tilde{A}^{-1} \check{A} \\ 0 & \text{id}_{n-2} \end{bmatrix} \begin{bmatrix} I \\ v \\ \hat{y}_{n-2} \end{bmatrix}. \end{aligned}$$

The composition of transformations (3.28) and (3.31) produces the following relation

$$(3.32) \quad \begin{aligned} \begin{bmatrix} \tilde{x} \\ \hat{x}_{n-2} \end{bmatrix} &= \begin{bmatrix} \text{id}_2 & 0 \\ \check{A}^t \check{A}^{-t} & \text{id}_{n-2} \end{bmatrix} \begin{bmatrix} \tilde{x}((J, \mathbf{v}) + \tilde{A}^{-1} \check{A} \hat{\mathbf{Y}}_{n-2}, \psi, \mathbf{u}) \\ \hat{\mathbf{X}}_{n-2} - \check{A}^t \check{A}^{-t}(\psi, \mathbf{u})^t \end{bmatrix}, \\ \begin{bmatrix} \tilde{Y} \\ \hat{Y}_{n-2} \end{bmatrix} &= \begin{bmatrix} \text{id}_2 & -\tilde{A}^{-1} \check{A} \\ 0 & \text{id}_{n-2} \end{bmatrix} \begin{bmatrix} \tilde{y}((J, \mathbf{v}) + \tilde{A}^{-1} \check{A} \hat{\mathbf{Y}}_{n-2}, \psi, \mathbf{u}) \\ \hat{\mathbf{Y}}_{n-2} \end{bmatrix}. \end{aligned}$$

Be aware of the fact that  $\tilde{x}^*$  denotes a closed curve with homological class  $[\tilde{x}_0] = (1, 0)$  (see Formula (3.27)), we have

$$\tilde{x}(I, \varphi + 1, u, v) = \tilde{x}(I, \varphi, u, v) + (1, 0), \quad \tilde{y}(I, \varphi + 1, u, v) = \tilde{y}(I, \varphi, u, v),$$

we get following from (3.32)

$$\begin{aligned} x(J, \psi + 1, \mathbf{u}, \mathbf{v}, \hat{\mathbf{X}}_{n-2}, \hat{\mathbf{Y}}_{n-2}) &= x(J, \psi, \mathbf{u}, \mathbf{v}, \hat{\mathbf{X}}_{n-2}, \hat{\mathbf{Y}}_{n-2}) + \mathbf{e}_1, \\ Y(J, \psi + 1, \mathbf{u}, \mathbf{v}, \hat{\mathbf{X}}_{n-2}, \hat{\mathbf{Y}}_{n-2}) &= Y(J, \psi, \mathbf{u}, \mathbf{v}, \hat{\mathbf{X}}_{n-2}, \hat{\mathbf{Y}}_{n-2}). \end{aligned}$$

For  $k \geq 3$  we also have

$$\begin{aligned} x(J, \psi, \mathbf{u}, \mathbf{v}, \hat{\mathbf{X}}_{n-2} + \hat{\mathbf{e}}_k, \hat{\mathbf{Y}}_{n-2}) &= x(J, \psi, \mathbf{u}, \mathbf{v}, \hat{\mathbf{X}}_{n-2}, \hat{\mathbf{Y}}_{n-2}) + \mathbf{e}_k, \\ Y(J, \psi, \mathbf{u}, \mathbf{v}, \hat{\mathbf{X}}_{n-2} + \hat{\mathbf{e}}_k, \hat{\mathbf{Y}}_{n-2}) &= Y(J, \psi, \mathbf{u}, \mathbf{v}, \hat{\mathbf{X}}_{n-2}, \hat{\mathbf{Y}}_{n-2}). \end{aligned}$$

where  $\mathbf{e}_k \in \mathbb{R}^n$  is a unit vector, whose  $k$ -th entry is equal to one, all other entries are equal to zero,  $\hat{\mathbf{e}}_k \in \mathbb{R}^{n-2}$  is defined in the same way. Consequently, the composition of the transformations (3.28) and (3.31) produces a map on  $T^*\mathbb{T}^n$ .

Next we restrict  $\mathfrak{S}^* \mathbf{H}_S$  to the NHIC  $\Pi$  to obtain a Hamiltonian  $\bar{\mathbf{H}}$  of one less degree of freedom. The restriction is done by setting  $u = v = 0$ , which implies

$$\mathbf{u} = u = 0, \quad \mathbf{v} + \langle \eta_2, \hat{\mathbf{Y}}_{n-2} \rangle = v = 0,$$

if we use  $\eta_i$  to denote the  $i$ -th row of the matrix  $\tilde{A}^{-1} \check{A}$  for  $i = 1, 2$ . We substitute two equations back to  $\mathfrak{S}^* \mathfrak{K}_\delta^* \mathbf{H}_S$  to eliminate  $\mathbf{u}, \mathbf{v}$  from our list of variables. The resulting

function denoted by  $\bar{H}$  is still Hamiltonian following from Theorem 3.2. The reduced system  $\bar{H}$  has the form

$$(3.33) \quad \begin{aligned} \bar{H} = & \bar{h}(J, \hat{Y}_{n-2}) + \bar{P}_I(J, \psi, \hat{Y}_{n-2}) \\ & + \delta \bar{R}_I(\psi, \hat{X}_{n-2}, J, \hat{Y}_{n-2}) + \varepsilon^\sigma \bar{R}_{II}(\psi, \hat{X}_{n-2}, J, \hat{Y}_{n-2}), \end{aligned}$$

where the main part  $\bar{h}$  takes the form

$$(3.34) \quad \bar{h} = \tilde{h}(J + \langle \eta_1, \hat{Y}_{n-2} \rangle) + \frac{1}{\sqrt{\varepsilon}} \langle \hat{\omega}_{n-2}, \hat{Y}_{n-2} \rangle + \frac{1}{2} \langle \hat{Y}_{n-2}, B \hat{Y}_{n-2} \rangle,$$

The perturbation term  $\bar{R}_I$  depends on the new action variable in somehow special way

$$\bar{R}_I(\psi, \hat{X}_{n-2}, J, \hat{Y}_{n-2}) = \bar{R}_I(\psi, \hat{X}_{n-2}, J + \langle \eta_1, \hat{Y}_{n-2} \rangle)$$

similarly for  $\bar{P}_I$ , and we have  $|\bar{R}_I|_{r-2} = O(1)$ ,  $|\bar{R}_{II}|_{r-2} = O(K^{2(2\tau+1)})$ .

To formalize the above procedure of order reduction, let us recall the operator  $\pi_i$  introduced at the beginning of Section 2. Denote by  $\pi_{-2}\mathfrak{S}$  the symplectic transformation obtained from  $\mathfrak{S}$  by applying  $\pi_{-2}$  to both the action and the angular variables, we have  $\pi_{-2}(\mathfrak{S}^{-1}) = (\pi_{-2}\mathfrak{S})^{-1}$ . One can verify that  $\bar{H} = (\pi_{-2}\mathfrak{S})^* k_\delta^* H_S$  due to  $u = v = 0$ .

**3.8. Frequency tracking and refinement.** Let us recall the transformation for the frequency shown in (2.2)

$$\omega_a \rightarrow M' \omega_a \rightarrow M'' M' \omega_a,$$

where  $M'$  is defined in (2.3) and  $M''$  is defined implicitly in Lemma 2.2.

$$\lambda\left(a, \frac{P}{Q}, \frac{p}{q}, \hat{\omega}_{n-3}^*\right)^t \xrightarrow{M'} \lambda\left(a, 0, \frac{d_0}{qQ}, \hat{\omega}_{n-3}^*\right)^t.$$

Assume at certain  $a = a''$  there is the second resonance  $\tilde{\mathbf{k}}'' \in \mathbb{Z}^n$ , i.e.  $\langle \tilde{\mathbf{k}}'', M' \omega_{a''} \rangle = 0$  by Lemma 2.2. For small  $\Delta a = a - a''$ , we have

$$(3.35) \quad M'' M' \omega_a = M'' \left( a'' + \Delta a, 0, \frac{d_0}{qQ}, \hat{\omega}_{n-3}^* \right)^t = \left( 0, 0, \hat{\omega}_{n-2} \right)^t + \mathbf{k}^\dagger \Delta a$$

where  $\mathbf{k}^\dagger$  is the first column of  $M''$ . Notice that the second entry of the frequency  $\omega = M'' M' \omega_a$  is zero since the second entry of  $\mathbf{k}^\dagger$  is zero (see Lemma 2.1).

Next, we apply the transformation  $S$ , which does not change the first two entries of the frequency vector, so the first two entries of the frequency vector  $SM'' M' \omega_a$  is  $(k_1^\dagger \Delta a, 0)$ , where  $k_1^\dagger$  is the first entry of  $\mathbf{k}^\dagger$ .

Recall Section 3.6, there we have chosen the homological class  $g = (1, 0) \in H_1(\mathbb{T}^2, \mathbb{Z})$  and the Hamiltonian system of one degree of freedom  $\tilde{h}(I)$  has periodic orbits achieving all frequency  $\nu(1, 0)$ . Therefore, the frequency line  $SM'' M' \omega_a$  lies in the image of the frequency map of the system  $\mathfrak{K}_\delta^* H_S = \mathfrak{K}_\delta^* \mathfrak{S}^{-1*} \mathfrak{M}''^{-1*} \mathfrak{M}^{-1*} H \circ \phi$  where  $H \circ \phi$  is in Proposition 2.2.

Next, we have undone the transformation  $\mathfrak{S}$ , so in the system  $\mathfrak{S}^* \mathfrak{K}_\delta^* H_S$  has the frequency line  $M'' M' \omega_a$  in (3.35). Finally, after the reduction of order during which we eliminate the second entry 0 from  $M'' M' \omega_a$ , so in a neighborhood of the frequency line  $\pi_{-2}(M'' M' \omega_a)$ , the Hamiltonian system  $\bar{H}$  is defined.

After one step of the reduction of order, we are handed with a Hamiltonian system of  $(n-1)$  degrees of freedom  $\bar{H}$ , for which we need to choose a new path to approximate the frequency line (2.2). We further undo the transformation  $\mathfrak{M}''$  to transform the system  $\bar{H}$  to  $\tilde{H} := (\pi_{-2} \mathfrak{M}'')^* \bar{H}$ , where  $\pi_{-2} \mathfrak{M}''$  is again to apply  $\pi_{-2}$  to both the

action and angular part of the symplectic transformation  $\mathfrak{M}''$ . The new system  $\bar{H} := (\pi_{-2}\mathfrak{M}'')^*\bar{H}$  is defined in a  $\mu$  neighborhood of the frequency line

$$(3.36) \quad \pi_{-2}M'\omega_a = (\pi_{-2}M''^{-1})(\pi_{-2}S^{-1})\pi_{-2}(SM''M'\omega_a) = \lambda_a\left(a, \frac{d_0}{qQ}, \hat{\omega}_{n-3}^*\right)^t.$$

To relocate the frequency line, notice the matrix  $M'$  does not change the fourth entry of  $\omega_a$ , we pick a rational number which is  $\mu$ -close to  $\omega_4^*$  denoted by  $\frac{\bar{p}}{\bar{q}}$  satisfying

$$(3.37) \quad \left| \frac{\bar{p}}{\bar{q}} - \omega_4^* \right| \leq \mu, \quad \text{g.c.d.}(\bar{p}, \bar{q}) = 1,$$

use the line segment of frequency  $\bar{\omega}_a := \bar{\lambda}_a(a, \frac{d_0}{qQ}, \frac{\bar{p}}{\bar{q}}, \hat{\omega}_{n-4}^*)^t \in \mathbb{R}^{n-1}$  to approximate the line segment  $\pi_{-2}M'\omega_a = \lambda_a(a, \frac{d_0}{qQ}, \hat{\omega}_{n-3}^*)^t$ .

This frequency vector has the form of  $\omega_a$  in Section 2.1 since  $\hat{\omega}_{n-4}^* \in \text{DC}(n-3, \alpha, \tau)$ . We set  $\frac{d}{qQ} = \frac{\bar{p}}{\bar{q}}$  satisfying  $\text{g.c.d.}(\bar{P}, \bar{Q}) = 1$  and consider some large  $\bar{K}(\geq K)$  to distinguish whether the second resonance is strong or not. Now we are in a situation completely parallel to Section 2.1 if we endow a *bar* to all the notations and get rid of one dimension in all the vectors. Again we encounter the situation of single and double resonances. The first resonance relation is given by  $\bar{\mathbf{k}}' = (0, \bar{Q}\bar{p}, -\bar{q}\bar{P}, \hat{\omega}_{n-4}^*)/\bar{d}$  before acted by  $\bar{M}''$  where  $\bar{d} = \text{g.c.d.}(\bar{q}\bar{P}, \bar{p}\bar{Q})$ .

As we vary  $a$ , double resonance may appear. Some of the double resonance points are new while some are inherited from that before the reduction of order. Next, we show that a strong double resonance vector before the reduction of order always gives rise to a strong double resonance after the reduction of order.

**Lemma 3.5.** *Each second resonance vector  $\mathbf{k}''$  with  $\langle \mathbf{k}'', \omega_{a''} \rangle = 0$  at some  $a''$  gives rise to a second resonant vector  $\bar{\mathbf{k}}'' := \pi_{-2}(\tilde{\mathbf{k}}'') := \pi_{-2}(\mathbf{k}''M'^{-1})$  for the reduced frequency vector  $\bar{\omega}_a$  satisfying  $\langle \bar{\mathbf{k}}'', \bar{\omega}_{\bar{a}''} \rangle = 0$ , where  $\bar{a}'' = a'' - \frac{\bar{k}_4}{\bar{k}_1}(\omega_4^* - \frac{\bar{p}}{\bar{q}})$ . The vector  $\bar{\mathbf{k}}''(\pi_{-2}M''^{-1}) = \mathbf{e}_1 = (1, 0, \dots, 0)$ .*

*Proof.* Consider at a double resonance point  $a''$  we have

$$\langle \mathbf{k}'', \omega_{a''} \rangle = 0$$

for some vector  $\mathbf{k}'' \in \mathbb{Z}^n \setminus \text{span}\{\mathbf{k}'\}$  and  $|\mathbf{k}''| \leq K$ . We have

$$\langle \mathbf{k}''M'^{-1}, M'\omega_{a''} \rangle = 0.$$

Next denote by  $\bar{\mathbf{k}}'' := \pi_{-2}(\mathbf{k}''M'^{-1})$ . Since we have changed  $\pi_{-2}M'\omega_a$  to  $\bar{\omega}_a$  by changing the Diophantine number  $\omega_4^*$  to a rational number  $\frac{\bar{p}}{\bar{q}}$ , suppose  $\bar{\mathbf{k}}'' = (\bar{k}_1, \bar{k}_3, \dots, \bar{k}_n)$ , then we have

$$\langle \bar{\mathbf{k}}'', \bar{\omega}_{a''} \rangle = -\bar{k}_4\left(\omega_4^* - \frac{\bar{p}}{\bar{q}}\right).$$

We then move  $a''$  slightly to  $\bar{a}'' = a'' - \frac{\bar{k}_4}{\bar{k}_1}(\omega_4^* - \frac{\bar{p}}{\bar{q}})$  so that we get  $\langle \bar{\mathbf{k}}'', \bar{\omega}_{\bar{a}''} \rangle = 0$ .

Since the first row of  $M''$  is  $\tilde{\mathbf{k}}''$ , one has  $\bar{\mathbf{k}}'' = \pi_{-2}\tilde{\mathbf{k}}'' = \mathbf{e}_1$ . □

The total number of double resonance points is finite depending only on  $\bar{K}$ . Most of these double resonances are weak. The total number of strong double resonances is finite and does not depend on  $\bar{K}$ .

As for the frequency line reduced in the case of double resonance, we only need to convert this approximation, by the map  $M''$ , to the coordinates under consideration.

We keep it in mind that the approximation of  $\omega_4^*$  by a rational number  $\frac{\bar{p}}{\bar{q}}$  shown in (3.37) applies to all frequency line segments.

**3.9. Normal form after one step of order reduction around strong double resonance.** The Hamiltonian with  $n$  degrees of freedom is reduced to a system with one less degree of freedom. To continue, we need to handle three possibilities when we are going to reduce the system further to two less degrees of freedom:

- (1) for the Hamiltonian (3.2) inherited from the case of single resonance, if it is still at single resonance, we apply Proposition 2.1;
- (2) for the Hamiltonian (3.2) inherited from the case of single resonance, if it is a newly born double resonance, we apply Proposition 2.2;
- (3) for the Hamiltonian inherited from the case of double resonance, it is always at double resonance due to Lemma 3.5, so we apply Proposition 2.2.

The resonance sub-manifolds can be defined again for the reduced system, denoted by  $\bar{\Sigma}', \bar{\Sigma}''$ . They are  $(n-2)$  and  $(n-3)$ -dimensional respectively.

A number  $\bar{\mu}$  is given by Lemma 2.3, depending on the size of the remainder  $\bar{\delta}$ , such that in the neighborhood  $B(\bar{\omega}_a, \bar{\mu})$ , either Proposition 2.1 or 2.2 is applicable. Recall we have a covering of the neighborhood  $B(\Sigma'', \varepsilon^\sigma)$  by balls of radius  $\varepsilon^\sigma$  centered at  $\Sigma''$ , which becomes a covering by balls of radius  $\varepsilon^{\sigma-1/2}$  in the homogenized system. We have that  $\bar{\Sigma}''$  is a codimension one sub-manifold in  $\Sigma''$ . The covering descends to a covering of a  $\varepsilon^\sigma$  neighborhood of  $\bar{\Sigma}''$  after the reduction of order. We consider only those balls intersecting  $B(\bar{\omega}_a, \bar{\mu})$  and discard the rest. In each of the remaining balls, we perform further KAM iteration and reduction of order. The balls are deformed by the linear symplectic transformation  $\mathfrak{M}''\mathfrak{M}'$  for strong double resonance and by  $\mathfrak{M}'$  for single or weak double resonance, but the norms of these transformations are bounded by a constant independent of  $\delta, \bar{\delta}, \varepsilon$ .

Next, we obtain a normal form in a neighborhood of a double resonance point that is succeeded from that before the reduction of order.

We define

$$(3.38) \quad \omega_a^2 = \bar{\lambda}_a \left( a, \frac{P}{Q}, \frac{p}{q}, \frac{\bar{p}}{\bar{q}}, \hat{\omega}_{n-4}^* \right)^t$$

as the frequency segment for the system (1.1), where the superscript 2 counts the current step of the reduction of order and approximation of frequency vector. We have that the frequency vector  $\bar{\omega}_a$  obtained in the previous section is related to  $\omega_a^2$  through

$$\pi_{-2}(M' \omega_a^2) = \bar{\omega}_a.$$

The frequency  $\omega_a^2$  has three resonance relations given by  $\mathbf{k}', \mathbf{k}''$  and  $\mathbf{k}'^2$ , where  $\mathbf{k}'^2 = (\pi_{+2}\bar{\mathbf{k}}')M'^{-1}$  for  $a = \bar{a}''$  given in Lemma 3.5.

Now we determine the triple resonance sub manifold as

$$\Sigma''' := \{y \mid \langle \mathbf{k}', \omega(y) \rangle = \langle \mathbf{k}'', \omega(y) \rangle = \langle \mathbf{k}'^2, \omega(y) \rangle = 0\} \subset \Sigma''.$$

We work on the Hamiltonian (2.19) in Proposition 2.2 with frequency  $\omega^* \in \Sigma'''$  of the form  $\omega_a^2$  for some  $a$ . We have the following normal form.

**Proposition 3.2.** *For all small  $\bar{\delta} > 0$ , there exists a symplectic transformation  $\bar{\phi}$  defined on  $(\varepsilon^{-1/2}\partial h^{-1}(B(\omega^*, \varepsilon^\sigma))) \times \mathbb{T}^n$ , which is  $O(\sqrt{\varepsilon}\bar{\delta}^{-(\tau+1)})$  close to identity in*

the  $C^{r-2}$  topology,  $|\bar{\phi} - \text{id}|_{r-2} = O(\sqrt{\varepsilon} \bar{\delta}^{-(\tau+1)})$ , sending the Hamiltonian  $H \circ \phi \in C^{r-2}$  of (2.19) to the following form provided  $\varepsilon$  is sufficiently small

$$(3.39) \quad \begin{aligned} H \circ \phi \circ \bar{\phi}(x, Y) = & \frac{1}{\varepsilon} h(Y^*) + \frac{1}{\sqrt{\varepsilon}} \langle \omega^*, Y \rangle + \frac{1}{2} \langle AY, Y \rangle + V(\langle \mathbf{k}', x \rangle, \langle \mathbf{k}'', x \rangle) \\ & + P_I(Y) + \delta \bar{V}(\langle \mathbf{k}', x \rangle, \langle \mathbf{k}'', x \rangle, \langle \mathbf{k}'^2, x \rangle) \\ & + \bar{\delta} \bar{\delta} \bar{R}_I(x) + \varepsilon^\sigma \bar{R}_{II}(x, Y), \end{aligned}$$

where the first five terms in the right-hand-side of (3.39) are the same as that in (2.19) of Proposition 3.2 with the same notations,  $\bar{V}$  consists of all the Fourier modes in  $\mathbf{R}$  in (2.19) of Proposition 3.2 in  $\text{span}_{\mathbb{Z}}\{\mathbf{k}', \mathbf{k}'', \mathbf{k}'^2\}$  and the remainder  $\bar{\delta} \bar{\delta} \bar{R}_I$  contains all the remaining Fourier modes in  $\delta \mathbf{R}_1$ . Moreover, we estimate the norms

$$|V|_r, |\bar{V}|_{r-2}, |\bar{R}_I|_{r-4}, |\bar{R}_{II}|_{r-4} \leq O(1), \quad \text{as } \varepsilon \ll \bar{\delta} \ll \delta \rightarrow 0.$$

*Proof.* The proof is almost the same as that of Proposition 3.2 so we only point out the difference. As  $\omega^*$  satisfies  $\langle \mathbf{k}', \omega^* \rangle = \langle \mathbf{k}'', \omega^* \rangle = 0$ , we get

$$\langle \mathbf{k}, \omega^* \rangle = \langle \mathbf{k} M'^{-1}, M' \omega^* \rangle = \langle \pi_{-2}(\mathbf{k} M'^{-1}), \pi_{-2}(M' \omega^*) \rangle, \quad \forall \mathbf{k} \in \mathbb{Z}^n,$$

since the second entry of  $M' \omega^*$  is zero. Due to the choice of the  $\bar{\mu}$  neighborhood, we have  $\pi_{-2}(M' \omega^*) \in B(\bar{\omega}_{\bar{a}''}, \bar{\mu})$ . So part (2) of Lemma 2.3 is applicable to  $\pi_{-2}(M' \omega^*)$ . Hence we get the estimate

$$|\langle \mathbf{k}, \omega^* \rangle| \geq \frac{\alpha \cdot \inf_a \lambda_a}{2^{\tau+1} (\bar{q} \bar{Q})^{\tau+1} (\|\bar{M}'\|_{\infty} \bar{K})^{2\tau+1}}$$

holds  $\forall \mathbf{k} \in \mathbb{Z}^n \setminus \text{span}\{\mathbf{k}', \mathbf{k}'', \mathbf{k}'^2\}$  with  $|\mathbf{k}| < \bar{K} = \bar{\delta}^{-1/2}$ . The proof of Proposition 3.2 goes through with this estimate.  $\square$

**3.10. Reduction of order around triple resonance.** In this section, we perform the reduction of order around the triple resonance.

**3.10.1. Linear symplectic transformation.** We have defined matrices  $M' \in SL(n, \mathbb{Z})$ . We introduce two matrices  $\mathbf{M}'_2, \mathbf{M}''_2 \in SL(n, \mathbb{Z})$  as follows

$$\mathbf{M}'_2 = \begin{bmatrix} \text{id}_2 & 0 & 0 & 0 \\ 0 & \frac{\bar{Q}\bar{P}}{d} & -\frac{\bar{q}\bar{P}}{d} & 0 \\ 0 & \bar{r} & \bar{s} & 0 \\ 0 & 0 & 0 & \text{id}_{n-4} \end{bmatrix} = \pi_{+2} \bar{M}', \quad \mathbf{M}''_2 = \begin{bmatrix} & \tilde{\mathbf{k}}''_2 & & \\ \tilde{0}_2 & \text{id}_2 & & 0_{2 \times (n-3)} \\ \hat{*}_{n-3} & 0_{(n-3) \times 2} & & *_{(n-3)^2} \end{bmatrix}$$

where the vector  $\tilde{\mathbf{k}}''_2 \in \mathbb{R}^n$ , first row of  $\mathbf{M}''_2$ , is defined from the relation  $\tilde{\mathbf{k}}''_2 \mathbf{M}'_2 M' = \mathbf{k}''$ ,  $\tilde{0}_2 = (0, 0)$ ,  $\hat{*}_{n-3}$  is an  $(n-3)$  dimensional vector and the  $*$  entries are defined as follows. In  $\mathbf{M}''_2$ , we remove the second and third columns and rows to get a matrix of  $(n-2) \times (n-2)$ . Then from the resulting first row, we find  $n-3$  integer vectors spanning unit volume with the first row. The matrix  $\mathbf{M}''_2$  determined in this way is in  $SL(n, \mathbb{Z})$ . We next define

$$M = \mathbf{M}''_2 \mathbf{M}'_2 M' \in SL(n, \mathbb{Z}),$$

By definition, the first three rows of  $M$  are  $\mathbf{k}'', \mathbf{k}'$  and  $\mathbf{k}'^2$  respectively. The matrix  $M$  induces a symplectic transformation

$$\mathfrak{M}: (x, Y) \mapsto (Mx, M^{-t}Y), \quad \omega_a = M\omega_a^2, \quad A = MAM^t,$$

where  $\omega_a$  has vanishing first three entries if  $a = \bar{a}''$  is such that  $\omega_{\bar{a}''}$  is at triple resonance. We denote  $\Delta a = a - \bar{a}''$ , so we get  $\omega_a = \omega_{\bar{a}''} + \Delta a e_1$ .

By the symplectic transformation  $\mathfrak{M}$ , one obtains the Hamiltonian defined on  $T^*\mathbb{T}^n$

$$\begin{aligned}
 (3.40) \quad H &:= \mathfrak{M}^{-1*}(\mathbf{H} \circ \phi \circ \bar{\phi}) \\
 &= \frac{1}{\varepsilon} h(Y^*) + \frac{1}{\sqrt{\varepsilon}} \langle \hat{\omega}_{\bar{a}'', n-3}, \hat{Y}_{n-3} \rangle + \frac{1}{2} \langle AY, Y \rangle + V(x_1, x_2) \\
 &\quad + \delta \bar{V}(x_1, x_2, x_3) + P_I(Y) + \bar{\delta} \delta R_I(x) + \varepsilon^\sigma R_{II}(x, Y),
 \end{aligned}$$

where  $P_I = \mathfrak{M}^{-1*} \mathbf{P}_I$ , similarly for  $R_I, R_{II}$ . The matrix  $M$  depends on  $\delta$  through  $\mathbf{k}'^2$  but independent of  $\bar{\delta}$ . The transformation  $\mathfrak{M}$  changes the norms of  $R_I, R_{II}$ . However, since we can choose  $\bar{\delta}, \varepsilon$  as small as we wish, the remainders can be arbitrarily small.

**3.10.2. Shear transformations.** We next introduce the shear transformation as we did in Lemma 3.1 to block diagonalize  $A$ . Let  $A, S_3 \in SL(n, \mathbb{R})$  be defined as follows

$$A = \begin{bmatrix} \tilde{A}_3 & \check{A}_3 \\ \check{A}_3^t & \hat{A}_3 \end{bmatrix}, \quad S_3 = \begin{bmatrix} \text{id}_3 & 0 \\ -\check{A}_3^t \tilde{A}_3^{-t} & \text{id}_{n-3} \end{bmatrix}$$

where  $\tilde{A}_3, \check{A}_3, \hat{A}_3$  are in  $\mathbb{R}^{3^2}, \mathbb{R}^{3 \times (n-3)}$  and  $\mathbb{R}^{(n-3)^2}$  respectively. With the shear matrix we introduce a transformation

$$\mathfrak{S}_3 : (x, Y) \mapsto (S_3 x, S_3^{-t} Y) := (\mathbf{x}, \mathbf{y}),$$

which transforms the Hamiltonian into the following system defined on  $T^*\mathbb{T}^3 \times \mathbb{R}^{2(n-3)}$

$$\begin{aligned}
 (3.41) \quad \mathbf{H}_{S_3} &:= \mathfrak{S}_3^{-1*} \mathfrak{M}^{-1*}(\mathbf{H} \circ \phi \circ \bar{\phi}) \\
 &= \frac{1}{\varepsilon} h(Y^*) + \left[ \frac{1}{2} \langle \tilde{A}_3 \tilde{\mathbf{y}}_3, \tilde{\mathbf{y}}_3 \rangle + V(\tilde{\mathbf{x}}) + \delta \bar{V}(\tilde{\mathbf{x}}_3) \right] \\
 &\quad + \frac{1}{\sqrt{\varepsilon}} \langle \hat{\omega}_{\bar{a}'', n-3}, \hat{\mathbf{y}}_{n-3} \rangle + \frac{1}{2} \langle B_3 \hat{\mathbf{y}}_{n-3}, \hat{\mathbf{y}}_{n-3} \rangle \\
 &\quad + (\bar{\delta} \delta \mathfrak{S}_3^{-1*} R_I(\mathbf{x}) + \varepsilon^\sigma \mathfrak{S}_3^{-1*} R_{II}(\mathbf{x}, \mathbf{y}) + P_I(S_3^t \mathbf{y})),
 \end{aligned}$$

where we denote  $B_3 = (\hat{A}_3 - \check{A}_3^t \tilde{A}_3^{-1} \check{A}_3)$  and  $\tilde{\mathbf{x}}_3 = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ ,  $\tilde{\mathbf{y}}_3 = (\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3)$ ,  $\tilde{\mathbf{x}} = (\mathbf{x}_1, \mathbf{x}_2)$ . The norms of the matrices  $B_3, S_3$  depends on  $\delta$  but not on  $\bar{\delta}$ . We next block diagonalize the quadratic form  $\langle \tilde{A}_3 \tilde{\mathbf{y}}_3, \tilde{\mathbf{y}}_3 \rangle$  by introducing one more shear transformation

$$S_2 = \begin{bmatrix} \text{id}_2 & 0 & 0 \\ -\mathbf{a}_3 \tilde{A}^{-1} & 1 & 0 \\ 0 & 0 & \text{id}_{n-3} \end{bmatrix},$$

where  $\mathbf{a}_3 = (a_{31}, a_{32}) \in \mathbb{R}^2$  is the vector formed by the entries of  $A$  on the third row to the left of the diagonal. We can verify that

$$(3.42) \quad S_2 \begin{bmatrix} \tilde{A}_3 & 0 \\ 0 & B_3 \end{bmatrix} S_2^t = \begin{bmatrix} \tilde{A} & 0 & 0 \\ 0 & b_3 & 0 \\ 0 & 0 & B_3 \end{bmatrix}$$

where  $b_3 = a_{33} - \mathbf{a}_3 \tilde{A}^{-1} \mathbf{a}_3^t$ . Denoting  $S = S_2 S_3$  and  $\mathfrak{S} : (\mathbf{x}, \mathbf{y}) \mapsto (S\mathbf{x}, S^{-t}\mathbf{y})$ , we get the following system defined on  $T^*\mathbb{T}^2 \times \mathbb{R}^{2(n-2)}$

$$\begin{aligned}
 (3.43) \quad \mathbf{H}_S &:= \mathfrak{S}^{-1*} \mathfrak{M}^{-1*}(\mathbf{H} \circ \phi \circ \bar{\phi}) \\
 &= \frac{1}{\varepsilon} h(Y^*) + \frac{1}{2} \langle \tilde{A} \tilde{\mathbf{y}}, \tilde{\mathbf{y}} \rangle + V(\tilde{\mathbf{x}}) + \frac{b_3}{2} \mathbf{y}_3^2 + \delta \bar{V}(S_2^{-1} \tilde{\mathbf{x}}_3) \\
 &\quad + \left[ \frac{1}{\sqrt{\varepsilon}} \langle \hat{\omega}_{\bar{a}'', n-3}, \hat{\mathbf{y}}_{n-3} \rangle + \frac{1}{2} \langle B_3 \hat{\mathbf{y}}_{n-3}, \hat{\mathbf{y}}_{n-3} \rangle \right] \\
 &\quad + P_I(S^t \mathbf{y}) + \mathfrak{S}^{-1*} (\bar{\delta} \delta R_I(\mathbf{x}) + \varepsilon^\sigma R_{II}(\mathbf{x}, \mathbf{y})).
 \end{aligned}$$



The frequency line now takes the form

$$(3.44) \quad \begin{aligned} SM\omega_a &= SM(\omega_{\bar{a}''} + \Delta a \mathbf{e}_1) \\ &= (0, 0, 0, \widehat{\omega_{\bar{a}''}_{n-3}}) + \Delta a \mathbf{k}^\dagger + \Delta a(0, 0, -s, \hat{0}_{n-3}), \end{aligned}$$

where  $\mathbf{k}^\dagger = (*, 0, 0, \hat{*}_{n-3})$  is the first column of  $\mathbf{M}_2''$ . Notice that the frequency line has 0 as the second entry.

**3.10.3. Reduction of order.** In the above system  $H_S$ , we apply Theorem 3.4 with homology class  $g = (1, 0)$  and find NHIC in the subsystem  $\tilde{\mathbf{G}} := \frac{1}{2}\langle \tilde{A}\tilde{y}, \tilde{y} \rangle + V(\tilde{x})$  since the current frequency line (3.44) has zero second entry. Restricted to the NHIC, the subsystem  $\tilde{\mathbf{G}}$  is reduced to a system of one degree of freedom denoted by  $\tilde{h}(I)$  in action-angle coordinates. We restrict to the region  $|I| > \gamma$  for some small number  $\gamma$  to be determined later.

Applying Theorem 3.2, one reduces the system to the following

$$(3.45) \quad \begin{aligned} \bar{H} &:= k_\delta^* \mathfrak{S}^{-1*} \mathfrak{M}^{-1*} (H \circ \phi \circ \bar{\phi}) \\ &= \varepsilon^{-1} h(Y^*) + \tilde{h}(I) + \frac{b_3}{2} y_3^2 + \delta \bar{Z}(I, \varphi, x_3) \\ &\quad + \left[ \varepsilon^{-1/2} \langle \hat{\omega}_{n-3}, \hat{y}_{n-3} \rangle + \frac{1}{2} \langle B_3 \hat{y}_{n-3}, \hat{y}_{n-3} \rangle \right] \\ &\quad + \bar{P}_I + (\delta \bar{R}_I(I, \varphi, \hat{x}_{n-2}) + \varepsilon^\sigma \bar{R}_{II}(I, \varphi, \hat{x}_{n-2}, \hat{y}_{n-2})), \end{aligned}$$

where  $\bar{Z} = k_\delta^*(S^{-1*})\bar{V}$  is obtained from  $(S^{-1*})\bar{V}$  by substituting  $x_1(I, \varphi), x_2(I, \varphi)$ . Similarly for  $\bar{P}_I, \bar{R}_I, \bar{R}_{II}$ .

**3.10.4. Undoing the shear transformation.** Now we get a subsystem  $\tilde{h}(I) + \frac{b_3}{2} y_3^2 + \delta \bar{Z}$  of two degrees of freedom. However, this system is not defined on  $T^*\mathbb{T}^2$  but on  $T^*\mathbb{T}^1 \times \mathbb{R}^2$ . To fix this issue, we can use the undo-shear method in Section 3.7 to get back a system defined on  $T^*\mathbb{T}^n$ . In this section, we only perform a partial undo-shear transformation to get a system defined on  $T^*\mathbb{T}^2 \times \mathbb{R}^{2(n-3)}$ .

We introduce the following undo-shear transformation

$$\underline{S} = \begin{bmatrix} 1 & 0 & 0 \\ s & 1 & 0 \\ 0 & 0 & \text{id}_{n-3} \end{bmatrix} \in SL(n-1, \mathbb{R}),$$

where  $s$  is the first entry of  $a_3 \tilde{A}^{-1}$  that is also the minus of the  $(3, 1)$  entry of  $S_2$ .

$$(3.46) \quad \begin{aligned} \underline{\mathfrak{S}} : (\varphi, x_3, \hat{x}_{n-3}; I, y_3, \hat{y}_{n-3}) &\mapsto (\underline{S}(\varphi, x_3, \hat{x}_{n-3}); \underline{S}^{-t}(I, y_3, \hat{y}_{n-3})) \\ &= (\varphi, s\varphi + x_3, \hat{x}_{n-3}; I - sy_3, y_3, \hat{y}_{n-3}) \\ &:= (\varphi, x_3; \hat{x}_{n-3}; J, y_3; \hat{y}_{n-3}). \end{aligned}$$

Under this transformation  $\underline{\mathfrak{S}}$ , the Hamiltonian becomes  $\bar{H}_{\underline{S}} := (\underline{\mathfrak{S}})^{-1*} \bar{H}$

$$(3.47) \quad \begin{aligned} \bar{H}_{\underline{S}} &:= \varepsilon^{-1} h(Y^*) + \left[ \tilde{h}(J + sy_3) + \frac{1}{2} b_3 y_3^2 + \delta \bar{Z}(J + sy_3, \varphi, x_3 - s\varphi) \right] \\ &\quad + \varepsilon^{-1/2} \langle \hat{\omega}_{n-3}, \hat{y}_{n-3} \rangle + \frac{1}{2} \langle B_3 \hat{y}_{n-3}, \hat{y}_{n-3} \rangle \\ &\quad + \bar{P}_I + \delta \bar{R}_I(J + sy_3, \varphi, x_3, \hat{x}_{n-3}) + \varepsilon^\sigma \bar{R}_{II}(\varphi, x_3; \hat{x}_{n-3}; J, y_3; \hat{y}_{n-3}). \end{aligned}$$

The same calculation as in (3.32) in Section 3.7 shows that the subsystem in the bracket is defined on  $T^*\mathbb{T}^2$ . Therefore, the system  $\bar{H}_{\underline{S}}$  is defined on  $T^*\mathbb{T}^2 \times \mathbb{R}^{2(n-3)}$ .

The frequency line now is obtained from (3.44) by removing the second entry 0 (first reduction) and replacing the third entry  $-s$  by 0 (undo-shear).

3.10.5. *Further reduction of order.* Again we need to distinguish the low and high energy regions. In the low energy region, the  $\bar{P}_1$  term is  $O(\sqrt{\varepsilon})$  hence is neglected.

Similar to what we met in Lemma 3.1, for the low energy region, we get a product of an integrable system of  $n-3$  degrees of freedom and a Tonelli system of two degrees of freedom

$$(3.48) \quad \tilde{G}_3(\varphi, x_3, J, y_3) = \tilde{h}(J + sy_3) + \frac{1}{2}b_3y_3^2 + \delta\bar{Z}(J + sy_3, \varphi, x_3 - s\varphi).$$

This system is perturbed by  $\bar{\delta}\delta\bar{R}_I + \varepsilon^\sigma\bar{R}_{II}$ . We will apply the procedure of order reduction to this system. Namely, we apply Theorem 3.4 with homology class  $g = (1, 0)$  to get a NHIC and restrict the system to the NHIC to get a system of one degree of freedom. We stick to the  $(1, 0)$  homology class since the current frequency line has zero second entry.

3.10.6. *Estimates of  $b_3$  and  $s$ .* We claim that

$$(3.49) \quad b_3 = \text{const}_b |\mathbf{k}'^2|^2, \quad s = \text{const}_s |\mathbf{k}'^2|,$$

where the constants are independent of  $\delta$ ,  $\text{const}_b > 0$  and  $\text{const}_s \in \mathbb{R}$ .

Recall the definitions of  $b_3$  (see (3.42))

$$b_3 = a_{33} - \mathbf{a}_3 \tilde{A}^{-1} \mathbf{a}_3^t$$

and  $s$  is the first entry of  $\mathbf{a}_3 \tilde{A}^{-1}$ . The  $(i, j)$ -th entry of  $A = MAM^t$  is  $m_i A m_j^t$  where  $m_i, m_j$  are the  $i$ -th and  $j$ -th rows of  $M$  respectively. Since the first three rows of  $M$  are  $\mathbf{k}'', \mathbf{k}', \mathbf{k}'^2$  respectively, we get that

$$b_3 = \mathbf{k}'^2 A (\mathbf{k}'^2)^t - (\mathbf{k}'^2 A \mathbf{K}^t) (\mathbf{K} A \mathbf{K}^t)^{-1} (\mathbf{K} A (\mathbf{k}'^2)^t),$$

and  $s$  is the first entry of  $\mathbf{k}'^2 A \mathbf{K}^t (\mathbf{K} A \mathbf{K}^t)^{-1}$ , where we denote by  $\mathbf{K}$  the matrix of  $2 \times n$  whose two rows are  $\mathbf{k}''$  and  $\mathbf{k}'$  respectively. Now  $s$  is estimated easily as  $\text{const} \cdot |\mathbf{k}'^2|$  since  $\mathbf{K} A \mathbf{K}^t (\mathbf{K} A \mathbf{K}^t)^{-1}$  does not depend on  $\delta$ .

We focus on  $b_3$  in the following. Since  $A$  is positive definite, we decompose  $A = CC^t$  for some  $C \in GL(n, \mathbb{R})$  and denote  $\mathbf{k}'^2 C := \mathbf{k}$ ,  $\mathbf{K} C = \mathbf{K}$ . This gives us

$$b_3 = \mathbf{k}(\mathbf{k}^t - \mathbf{K}^t (\mathbf{K} \mathbf{K}^t)^{-1} \mathbf{K} \mathbf{k}^t).$$

Now, we recall the Gauss least square method. The equation  $\mathbf{K}^t \mathbf{x} = \mathbf{k}^t$ , though in general not solvable for  $\mathbf{x} \in \mathbb{R}^2$ , we can seek for a least square solution given by  $\mathbf{x}_{ls} = (\mathbf{K} \mathbf{K}^t)^{-1} \mathbf{K} \mathbf{k}^t$ , which has geometric interpretation as follows. The vector  $\mathbf{K} \mathbf{x}_{ls} = \mathbf{K} (\mathbf{K} \mathbf{K}^t)^{-1} \mathbf{K} \mathbf{k}^t$  is the projection of  $\mathbf{k}$  to the linear space spanned by the column vectors of  $\mathbf{K}^t$ . Hence  $(\mathbf{k}^t - \mathbf{K}^t (\mathbf{K} \mathbf{K}^t)^{-1} \mathbf{K} \mathbf{k}^t)$  is the projection of  $\mathbf{k}$  to the orthogonal complement of the linear space spanned by the column vectors of  $\mathbf{K}^t$ . We see from the construction of the vectors  $\mathbf{k}'', \mathbf{k}', \mathbf{k}'^2$  that  $\mathbf{k}'^2$  forms a nonzero angle with the plane  $\text{span}\{\mathbf{k}', \mathbf{k}''\}$  independent of  $\delta$ , since as  $\mu \rightarrow 0$  one has

$$\bar{\mathbf{k}}' = (0, \bar{Q}\bar{p}/\bar{d}, -\bar{q}\bar{P}/\bar{d}, \hat{0}_{n-4}) \parallel (0, 1, -\frac{\bar{q}\bar{P}}{\bar{Q}\bar{p}}, \hat{0}_{n-3}) \rightarrow (0, 1, -\omega_4^* \bar{Q}/\bar{P}, \hat{0}_{n-3}),$$

from which  $\mathbf{k}'^2 = (\pi_{+2} \bar{\mathbf{k}}') M'^{-1}$  is defined and the matrices  $\mathbf{K}, M', A$  do not depend on  $\delta$ . This linear independence relation is preserved by the linear transformation  $C$ . Hence we get that  $b_3 = c |\mathbf{k}'^2|^2$  for some constant  $c > 0$  and independent of  $\delta$ .

**3.11. Further genericity conditions.** In this section, we talk about the genericity conditions that are needed during the second step of reduction of order.

**3.11.1. Genericity conditions in the regime free of strong double resonances.** For the frequency vector  $\omega_a^2 = \bar{\lambda}_a(a, \frac{P}{Q}, \frac{p}{q}, \frac{\bar{p}}{q}, \hat{\omega}_{n-4}^*)$  associated to the original system (1.1), suppose there are only two resonance vectors  $\mathbf{k}', \mathbf{k}'^2$ .

Let  $Z_2(\langle \mathbf{k}', x \rangle, \langle \mathbf{k}'^2, x \rangle, y) := Z(\langle \mathbf{k}', x \rangle, y) + \delta \bar{Z}(\langle \mathbf{k}', x \rangle, \langle \mathbf{k}'^2, x \rangle, y)$  be formed by the Fourier modes of  $P(x, y)$  in the  $\text{span}_{\mathbb{Z}}\{\mathbf{k}', \mathbf{k}'^2\}$ , where  $Z'$  agrees with that in **(H1)** in Section 3.2. We can introduce a matrix  $M \in SL(n, \mathbb{Z})$  whose second and third rows are  $\mathbf{k}', \mathbf{k}'^2$  to induce the following symplectic transformation

$$(x, y) \mapsto (Mx, M^{-t}y) := (\mathbf{x}, \mathbf{y})$$

In the new coordinates, the function

$$Z_2(\langle \mathbf{k}', x \rangle, \langle \mathbf{k}'^2, x \rangle, y) \mapsto Z'_2(\mathbf{x}_2, \mathbf{x}_3, M^t \mathbf{y}) = Z'(\mathbf{x}_2, M^t \mathbf{y}) + \delta \bar{Z}'(\mathbf{x}_2, \mathbf{x}_3, M^t \mathbf{y})$$

We introduce the following non degeneracy condition.

**(H1)<sub>2</sub>** *Uniformly for all  $y \in \Gamma'_{a,2} := \partial h^{-1}(\omega_a^2)$ , the function  $Z'_2(\mathbf{x}_2, \mathbf{x}_3, y)$  has unique non degenerate global maximum, namely, the Hessian of  $Z'_2(\mathbf{x}_2^*(y), \mathbf{x}_3^*(y), y)$  in  $(\mathbf{x}_2, \mathbf{x}_3)$  is negative definite, where  $(\mathbf{x}_2^*(y), \mathbf{x}_3^*(y))$  is the global maximum, with finitely many exceptions for which there are two non degenerate global maxima.*

The genericity of this condition **(H1)<sub>2</sub>** is guaranteed by the following result.

**Theorem 3.5** ([CZ2]). *Assume  $M$  is a closed manifold with finite dimensions,  $F_\zeta \in C^r(M, \mathbb{R})$  with  $r \geq 4 \forall \zeta \in [\zeta_0, \zeta_1]$  and  $F_\zeta$  is Lipschitz in the parameter  $\zeta$ . Then, there exists an open-dense set  $\mathfrak{V} \subset C^r(M, \mathbb{R})$  so that for each  $V \in \mathfrak{V}$ , it holds simultaneously for all  $\zeta \in [\zeta_0, \zeta_1]$  that the minimum of  $F_\zeta + V$  is non-degenerate. In fact, given  $V \in \mathfrak{V}$  there are finitely many  $\zeta_i \in [\zeta_0, \zeta_1]$  such that  $F_\zeta + V$  has only one global minimal (maximal) point for  $\zeta \neq \zeta_i$  and has two global minimal (maximal) point if  $\zeta = \zeta_i$ .*

Next, we show that the genericity condition **(H1)<sub>2</sub>** that is imposed on the original Hamiltonian (1.1) descends to a version of **(H1)** imposed on the reduced system of  $n-1$  degrees of freedom. As  $\Gamma'_a$  lies in a  $O_{\delta \rightarrow 0}(\mu)$  neighborhood of  $\Gamma'_{a,2}$ , the non-degeneracy condition **(H1)** still holds on the curve  $\Gamma'_{a,2}$  by the implicit function theorem. For each  $M^t \mathbf{y} \in \Gamma'_{a,2}$ , suppose the global maximum of  $Z'$  is achieved at  $\mathbf{x}_2^*(y)$  (if there are two, we choose either of them). We substitute this to the expression of  $Z'_2$  to get  $\delta \bar{Z}'(\mathbf{x}_2^*(y), \mathbf{x}_3, M^t \mathbf{y})$  plus a function of  $y$  only. For the resulting function of  $\mathbf{x}_3, y$ , we can apply **(H1)** to show that the global maximum with respect to  $\mathbf{x}_3$  are non-degenerate for all  $M^t \mathbf{y} \in \Gamma'_{a,2}$  generically. Suppose this global maximum is achieved at the point  $\mathbf{x}_3^*(y)$ . However, the point  $(\mathbf{x}_2^*(y), \mathbf{x}_3^*(y))$  does not necessarily achieve the global max of  $Z'_2$ . Instead, the implicit function theorem implies that the global maximum of  $Z'_2$  is achieved at a point  $(\mathbf{x}_2^*(y), \mathbf{x}_3^*(y))$  which lies in a  $O(\delta)$  neighborhood of  $(\mathbf{x}_2^*(y), \mathbf{x}_3^*(y))$ . Indeed we first see that  $|\mathbf{x}_2^*(y) - \mathbf{x}_2^*(y)| = O(\delta)$  since  $Z'_2$  is a  $\delta$  perturbation of  $Z'$ , next, if we modify  $\mathbf{x}_2^*(y)$  by  $O(\delta)$  in  $\bar{Z}'(\mathbf{x}_2^*(y), \mathbf{x}_3, M^t \mathbf{y})$ , the non-degeneracy implies that the global maximum is achieved at a point that is  $O(\delta)$  close to  $\mathbf{x}_3^*(y)$ .

Similar to Section 3.2, the global maximum of  $Z'_2$  gives rise to a NHIC of dimension  $2(n-2)$  whose normal Lyapunov exponents are  $\pm O(1)$  and  $\pm O(\sqrt{\delta})$  as  $\delta, \varepsilon \rightarrow 0$ . We also get such a NHIC by applying the procedure of order reduction to the system (3.2) restricted to the NHIC of dimension  $2(n-1)$ . The two NHIC stays  $O(\delta)$  close to each other by the previous paragraph, hence coincide by the local uniqueness.

3.11.2. *The genericity condition on new strong double resonance.* Suppose for some  $a$ , there exist three resonance vectors  $\{\mathbf{k}', \mathbf{k}'', \mathbf{k}'^2\}$  to  $\omega_a^2$  associated to the original system (1.1), where  $\mathbf{k}', \mathbf{k}''$  give rise to a strong double resonance for  $\omega_a$  before the first step of reduction of order.

We need to impose a new genericity condition on the original Hamiltonian system (1.1) to show that the reduced system (3.48) satisfies Theorem 3.4. Here we do not talk about the analogues of **(H2.1, H2.2)** since the spectral gap condition therein is satisfied automatically in the following steps of reduction of orders and we stay away from the strong double resonance by a distance  $\gamma$  (see Remark 3.2). Consider the bracketed system in (3.41)

$$\mathbf{H}_3 = \left[ \frac{1}{2} \langle \tilde{A}_3 \tilde{y}_3, \tilde{y}_3 \rangle + V(\tilde{\mathbf{x}}) + \delta \bar{V}(\tilde{\mathbf{x}}_3) \right], \quad (\tilde{\mathbf{x}}_3, \tilde{y}_3) \in T^* \mathbb{T}^3,$$

which is obtained from the system (1.1) by homogenization, picking out Fourier modes in  $\text{span}_{\mathbb{Z}}\{\mathbf{k}', \mathbf{k}'', \mathbf{k}'^2\}$  and a shear linear transformation which is identity applied to  $\tilde{\mathbf{x}}_3$ .

**(H2)<sub>2</sub>** Assume  $V \in C^r(\mathbb{T}^2, \mathbb{R})$  ( $r \geq 5$ ) satisfies **(H2)**. For generic  $\bar{V} \in C^r(\mathbb{T}^3, \mathbb{R})$  and small enough  $\delta$ , the action minimizing invariant measures with rotation vector  $\nu(1, 0, 0)$ ,  $|\nu| \in [\gamma, \infty)$ ,  $\forall \gamma > 0$ , are supported on one hyperbolic periodic orbits except for finitely many  $\nu$ 's with two hyperbolic periodic orbits. These periodic orbits makes up finitely many pieces NHICs.

*Proof.* The proof of the genericity of the above hypothesis is almost the same as that of [CZ1]. Here we sketch the main steps and point out the necessary modifications.

When  $\delta \bar{V} = 0$ , the Hamiltonian  $\mathbf{H}_3$  is free of  $\mathbf{x}_3$ . Therefore, one can treat it as a system with two degrees of freedom. It follows from [CZ1] that there are finitely many pieces of NHIC extending to  $\nu \rightarrow \infty$ . For any rotation vector of the form  $(*, *, 0)$  the minimal measure is supported on a manifold which is the product of these NHICs with the cylinder  $\mathbb{T} \times \mathbb{R} \ni (\mathbf{x}_3, y_3)$ . This manifold survives small perturbation  $\delta \bar{V}$ , due to the theorem of normally hyperbolic invariant manifold. By the upper semi-continuity of Mañé set with respect to small perturbation and the normal hyperbolicity one can see that the minimal measure with rotation vector of the form  $(*, *, 0)$  is still supported on the NHIMs (four-dimensional cylinders) provided  $\delta \bar{V}$  is small.

Restricted on this four-dimensional cylinder we obtain a reduced Hamiltonian with two degrees of freedom. Consequently, each minimal measure with resonant rotation vector must be supported on periodic orbits. Back to the original system, the minimal measure with rotation vector  $\nu(1, 0, 0)$  is supported on periodic orbit. So, we will do

*Step 1.* By the standard energetic reduction, we reduce the system  $\mathbf{H}_3$  to a system of 2.5 degrees of freedom by choosing  $y_1$  as the new Hamiltonian and  $-\mathbf{x}_1$  as the new time. Write the action functional defined on the space of loops with homology class  $(1, 0, 0)$  and further restrict to a function  $F_E(\mathbf{x}_2, \mathbf{x}_3)$  smooth in  $\mathbf{x}_2, \mathbf{x}_3$  and  $E$  by minimizing the action among curves with fixed initial and final points  $(\mathbf{x}_2, \mathbf{x}_3)$ . The global minimum of  $F_E(\mathbf{x}_2, \mathbf{x}_3)$  gives rise to the periodic orbits supporting the Mather set. We next show the generic non-degeneracy of the global minimum of  $F_E(\mathbf{x}_2, \mathbf{x}_3)$  for  $E$  in an interval  $[E_0, \infty)$  corresponding to the interval of  $\nu$ .

*Step 2.* The union of periodic orbits make up a cylinder. We pick a section of the periodic orbits with energy in  $[E_* - d, E_* + d]$  for suitably small  $d$ . Without loss of generality, we suppose the section is given by the  $\mathbf{x}_1, \mathbf{x}_3$  torus.

*Step 3.* One way to construct the perturbation is given in [CZ1], which has straightforward higher dimensional generalization. The idea is to straighten the flow in a neighborhood of the periodic orbit and construct the perturbation in such a way that it is the linear combination of the first several Fourier modes in the disk transversal to the periodic orbit. The advantage of the construction in [CZ1] is that the potential added to the Lagrangian is also added to the action function explicitly as a function of  $x_2, x_3$ .

*Step 4.* The next and crucial step is to use the smooth dependence on parameter  $E$  to show that the non-degeneracy of periodic orbits holds for parameters in a interval  $(E_* - d, E_* + d)$  for suitably small  $d$ . This is an application of Theorem 3.5. Then by taking an open cover of  $[E_0, E_1]$  for some large  $E_1$ , the compactness of the interval  $[E_0, E_1]$  gives that only finitely many perturbation is needed, which gives the genericity.

*Step 5.* Since the Hamiltonian  $H_3$  restricts to a system (3.48) of two degrees of freedom and the minimizing periodic orbits are also minimizing periodic orbits for the reduced system (3.48), we get that the periodic orbits are also nondegenerate for the reduced system (3.48) (a non degenerate positive definite matrix remains non degenerate when restricted to a linear subspace). Nondegeneracy implies hyperbolicity of periodic orbits following from the same argument as [CZ1] or the proof of Lemma 3.2 applied to the reduced system (3.48).

*Step 6.* To show the uniform hyperbolicity for the energy intervals  $[E_1, \infty)$  for large  $E_1$ , we apply the argument of Lemma 3.2 to the reduced system (3.48) to get that the normal hyperbolicity is determined by the hyperbolic fixed point of the system  $\frac{b_3}{2}y_3^2 + \delta[\bar{Z}](x_3)$  where  $[\bar{Z}](x_3) = \int_{\mathbb{T}^1} \bar{Z}(\infty, \varphi, x_3 - s\varphi) d\varphi$ . To see that  $\bar{Z}(\infty, \cdot, \cdot)$  is not singular, we recall that  $\bar{Z}$  is obtained from  $\bar{V}$  by restricting to the NHIC parametrized by  $(I, \varphi)$  and formed by action minimizing periodic orbits of the system (3.13) with homology class  $(1, 0)$ . For high enough energy level, i.e.  $|I|$  large enough, the action minimizing periodic orbit of the system (3.13) approaches the circle  $(y_1, x_1, y_2 = 0, x_2 = \arg\max \int V(\tilde{x}) dx_1)$  in the  $C^1$  norm (See Proposition 3.1 of [C15a]), hence  $(I(\tilde{x}, \tilde{y}), \varphi(\tilde{x}, \tilde{y})) \rightarrow (y_1, x_1)$  in the  $C^1$  norm. Since  $\bar{V}$  does not depend on  $y_1$ , we get  $\bar{Z}$ 's dependence on  $I$  is weaker and weaker as  $|I|$  gets larger.  $\square$

#### 4. A HIERARCHY OF CANONICAL TRANSFORMATIONS

In this section, we summarize the above algorithm of the reduction of order inductively. The contents in this section are purely algebraic without analysis or dynamics.

**4.1. Induction at the single and weak double resonances.** We are given a vector  $\hat{\omega}_{n-3}^* = (\omega_4^*, \dots, \omega_n^*) \in \text{DC}(n-2, \alpha, \tau)$  and use the notation

$$\hat{\omega}_{n-j}^* := (\omega_{j+1}^*, \dots, \omega_n^*) \in \text{DC}(n-j+1, \alpha, \tau), \text{ for } j = 3, 4, \dots, n-1.$$

Suppose we start with a Hamiltonian system with  $j = 0, 1, \dots, n-3$ ,

$$(4.1) \quad H_j(\mathbf{x}_j, \mathbf{y}_j) = \frac{1}{\sqrt{\varepsilon}} \langle \omega_j^*, \mathbf{y}_j \rangle + \frac{1}{2} \langle A_j \mathbf{y}_j, \mathbf{y}_j \rangle + \delta_j R_{I,j}(\mathbf{x}_j) + \varepsilon^{\frac{r}{2(r+2)}} R_{II,j}(\mathbf{x}_j, \mathbf{y}_j),$$

defined in a ball of radius 1 centered at 0 with

$$(1) \quad \omega^{j,*} \text{ satisfies } |\langle \mathbf{k}, \omega^{j,*} \rangle| \geq \varepsilon^{\frac{2}{r+2}} \delta_j^{-1/2} \text{ for } |\mathbf{k}| < K_j, \mathbf{k} \in \mathbb{Z}^n, \text{ and } K_j = \delta_j^{-1/2}$$

- (2)  $\omega^{j,*}$  lies in a  $\mu_j$  neighborhood of the frequency line

$$\omega_a^j = \lambda_{a,j} \left( a, \frac{P_{j+2}}{Q_{j+2}}, \frac{p_{j+3}}{q_{j+3}}, \hat{\omega}_{n-j-3}^* \right)^t \in \mathbb{R}^{n-j}$$

- (3) and the constant  $\mu_j$  is determined from Lemma 2.3 by a constant  $\delta$  that can be made as small as we wish. Denote this  $\delta$  by  $\delta_j$ .  
 (4) We adopt the following convention of labeling entries of a vector

$$\mathbf{v}_j = (v_1, v_{j+2}, v_{j+3}, \dots, v_n)$$

for any vector  $\mathbf{v}_j \in \mathbb{R}^{n-j}$ , which may represent  $\mathbf{x}_j, \mathbf{y}_j, \omega_a^j$ , etc.,

- (5) when  $j = 0$ ,  $\frac{P_{j+2}}{Q_{j+1}} = \frac{P}{Q}$  and  $\frac{p_{j+3}}{q_{j+3}} = \frac{p}{q}$ . Those with  $j > 0$  are determined inductively in the following. The torus  $\mathbb{T}_0^n$  is the standard torus  $\mathbb{R}^n/\mathbb{Z}^n$  and that for  $j > 0$  is also determined inductively in the following.  
 (6)  $A_j$  is positive definite and is independent of  $\delta_j$ .

As  $a$  varies in an interval, we always have a first resonance vector

$$(4.2) \quad \mathbf{k}'^j = \left( 0, \frac{Q_{j+2}P_{j+3}}{d_j}, -\frac{q_{j+3}P_{j+2}}{d_j}, \hat{0}_{n-j-3} \right) \in \mathbb{Z}^{n-j},$$

where  $d_j = \text{g.c.d.}\{q_{j+3}P_{j+2}, p_{j+3}Q_{j+2}\}$ . For some  $a$ , we have a second resonant vector denoted by  $\mathbf{k}''^j$ . We apply Proposition 2.1 in the case of single resonance to get the normal form for arbitrarily small  $\delta_{j+1}$

$$(4.3) \quad \begin{aligned} H_j(\mathbf{x}_j, \mathbf{y}_j) &= \frac{1}{\sqrt{\varepsilon}} \langle \omega_j^*, \mathbf{y}_j \rangle + \frac{1}{2} \langle A_j \mathbf{y}_j, \mathbf{y}_j \rangle + \delta_j V_j(\langle \mathbf{k}'^j, \mathbf{x}_j \rangle) \\ &\quad + \delta_{j+1} R_{I,j+1}(\mathbf{x}_j) + \varepsilon_j^{\frac{r}{2(r+2)}} R_{II,j+1}(\mathbf{x}_j, \mathbf{y}_j). \end{aligned}$$

This normal form is defined in a ball of radius 1 centered at  $\mathbf{y}_j^*$  whose corresponding frequency in  $B(\omega_a^j, \mu_{j+1})$ , where  $\mu_{j+1} (\ll \mu_j)$  is determined from  $\delta_{j+1}$  by applying Lemma 2.3. We introduce  $\delta_j$  as the threshold to distinguish the strong and weak double resonance in Section 3.3.

We also introduce the frequency vector

$$(4.4) \quad \omega_a^j = \lambda_{a,j} \left( a, \frac{P_2}{Q_2}, \frac{p_3}{q_3}, \dots, \frac{p_{j+3}}{q_{j+3}}, \hat{\omega}_{n-j-3}^* \right), \quad j = 0, 1, \dots, n-3$$

associated to the original system (1.1) and related to the frequency  $\omega_a^j$

$$\omega_a^j = M_0'^{-1} \left( \pi_{+2} M_1'^{-1} \dots \pi_{+2} \left( M_{j-1}'^{-1} \left( \pi_{+2} \omega_a^j \right) \right) \right).$$

The integer vector  $\mathbf{k}'^j \in \mathbb{Z}^{n-j}$  corresponds to integer vectors  $\mathbf{k}^j \in \mathbb{Z}^n$  that is perpendicular to  $\omega_a^j$ . We introduce a function  $Z_j(x, y)$  consisting of those Fourier modes of  $P(x, y)$  in  $\text{span}_{\mathbb{Z}}\{\mathbf{k}^1, \dots, \mathbf{k}^j\}$ . We next introduce a function  $Z'_j(\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_{j+1}, y)$  which is obtained from  $Z_j$  by transforming the  $x$  variables by a matrix in  $SL(n, \mathbb{Z})$  whose  $(j+1)$ -th row is  $\mathbf{k}^j$ . The genericity condition that we need is the following, guaranteed by Theorem 3.5.

**(H1)<sub>j</sub>**: *Uniformly for all  $y \in \Gamma'_{a,j} := \partial h^{-1}(\omega_a^j)$ , the function  $Z'_j(\mathbf{x}_2, \dots, \mathbf{x}_{j+1}, y)$  has unique non degenerate global maximum, i.e. the Hessian of  $Z'_j$  in  $(\mathbf{x}_2, \dots, \mathbf{x}_{j+1})$  is negative definite at  $(\mathbf{x}_2^*(y), \dots, \mathbf{x}_{j+1}^*(y))$  which is the global maxima point, with finitely many exceptions for which there are two non-degenerate global maxima.*

We get a unimodular matrix  $M'_j \in \mathbb{Z}^{(n-j)^2}$  such that

$$(4.5) \quad M'_j \omega_a^j = \left(a, 0, \frac{P_{j+3}}{Q_{j+3}}, \hat{\omega}_{n-j-3}^*\right), \quad M'_j = \begin{bmatrix} 1 & 0 & 0 \\ 0 & (\tilde{M}_j)_{2 \times 2} & 0 \\ 0 & 0 & \text{id}_{n-j-3} \end{bmatrix}.$$

The matrix  $\tilde{M}_j$  and the rational number  $\frac{P_{j+3}}{Q_{j+3}} = \frac{d_j}{Q_{j+2}q_{j+3}}$  are defined inductively as follows

$$\tilde{M}_j \begin{bmatrix} \frac{P_{j+2}}{Q_{j+2}} \\ \frac{Q_{j+2}}{p_{j+3}} \\ \frac{p_{j+3}}{q_{j+3}} \end{bmatrix} := \begin{bmatrix} \frac{Q_{j+2}p_{j+3}}{d_j} & -\frac{P_{j+2}q_{j+3}}{d_j} \\ r_j & s_j \end{bmatrix} \begin{bmatrix} \frac{P_{j+2}}{Q_{j+2}} \\ \frac{Q_{j+2}}{p_{j+3}} \\ \frac{p_{j+3}}{q_{j+3}} \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{P_{j+3}}{Q_{j+3}} \end{bmatrix}$$

where  $d_j = \text{g.c.d.} \{Q_{j+2}p_{j+3}, P_{j+2}q_{j+3}\}$  and  $r_j, s_j$  are integers such that

$$s_j(Q_{j+2}p_{j+3}) + r_j(P_{j+2}q_{j+3}) = d_j$$

which are determined uniquely by the Euclidean algorithm.

Next we denote the linear symplectic transformation

$$\mathfrak{M}'_j : (\mathbf{x}_j, \mathbf{y}_j) \mapsto (M'_j \mathbf{x}_j, M_j'^{-t} \mathbf{y}_j)$$

$$\pi_{-2} \mathfrak{M}'_j : (\mathbf{x}_j, \mathbf{y}_j) \mapsto (\pi_{-2}(M'_j \mathbf{x}_j), \pi_{-2}((M'_j)^{-t} \mathbf{y}_j)) := (\mathbf{x}_{j+1}, \mathbf{y}_{j+1})$$

so that each of  $\mathbf{x}_{j+1}$  and  $\mathbf{y}_{j+1}$  has  $n - (j + 1)$  components. After the reduction of order done in Section 3.2, we get a Hamiltonian system with the same form as (4.1) while the subscript  $j \rightarrow j + 1$ . We choose any rational number  $\frac{p_{j+4}}{q_{j+4}}$  satisfying

$$(4.6) \quad \left| \frac{p_{j+4}}{q_{j+4}} - \omega_{j+4}^* \right| < \mu_{j+1},$$

and denote the frequency vector

$$(4.7) \quad \omega_a^{j+1} = \lambda_{a,j+1} \left( a, \frac{P_{j+3}}{Q_{j+3}}, \frac{p_{j+4}}{q_{j+4}}, \hat{\omega}_{n-j-4}^* \right) \in \mathbb{R}^{n-(j+1)}.$$

Now we have complete the induction from step  $j$  to the step  $j + 1$  in the case of single resonance and weak double resonance.

**Proposition 4.1.** *Assume  $(\mathbf{H1})_j$  for the Hamiltonian system (1.1) along all the frequency segments  $\omega_a^j$  at single resonance. We repeat the above procedure for  $n - 2$  steps to get a frequency vector*

$$(4.8) \quad \omega_a^\# = \lambda_a^\# \left( a, \frac{p_2}{q_2}, \frac{p_3}{q_3}, \dots, \frac{p_n}{q_n} \right) = M_0'^{-1} (\pi_{+2} M_1'^{-1} \dots M_{n-4}'^{-1} (\pi_{+2} \omega_a^{n-3}))$$

where  $\frac{p_2}{q_2} = \frac{P}{Q}$ ,  $\frac{p_3}{q_3} = \frac{p}{q}$ ,  $\lambda_a^\#$  is such that  $\alpha(\mathcal{L}_\beta(\omega_a^\#)) = E$ , and we determine the list of  $\frac{p_j}{q_j}$  and a hierarchy of size of neighborhoods  $\mu_0 \gg \mu_1 \gg \dots \gg \mu_{j-3}$  in the following way:

- (1) Once  $\mu_j$  is chosen, then  $\frac{p_{j+3}}{q_{j+3}}$  is chosen to be any rational number satisfying

$$\left| \frac{p_{j+3}}{q_{j+3}} - \omega_{j+3}^{*i} \right| < \mu_j, \quad j = 0, 1, \dots, n-3.$$

- (2) Once  $\frac{p_{j+3}}{q_{j+3}}$  is chosen, we first choose  $\delta_{j+1} < \Delta(\mathbf{k}^j)$  applying Lemma 3.3, where  $\mathbf{k}^j$  is determined by  $\frac{p_2}{q_2}, \dots, \frac{p_{j+3}}{q_{j+3}}$  inductively through (4.2), then determine  $\mu_{j+1}$  from  $\delta_{j+1}$  applying Lemma 2.3 with  $K = \delta_{j+1}^{-1/2}$ .

The Hamiltonian system is reduced to a system of two degrees of freedom as follows

$$(4.9) \quad H_{n-2}(x, y) = \frac{1}{\sqrt{\varepsilon}} \omega y_2 + \frac{1}{2} \langle A(y_1, y_2), (y_1, y_2) \rangle + \delta_{n-2} R_I(x_1, x_2) + \varepsilon^{\frac{r}{2(r+2)}} R_{II}(x, y),$$

where  $A \in \mathbb{R}^{2 \times 2}$  is positive definite, the number  $\omega \neq 0$  is independent of  $\varepsilon, \delta_{n-2}$ ,  $|R_I|_{C^r}, |R_{II}|_{C^{r-2j}}$  are bounded as  $\delta_{n-2}, \varepsilon \rightarrow 0$  and  $\delta_{n-2}$  is as small as we wish.

In the resulting system (4.9), we will later perform a standard energetic reduction to reduce it to a system of 1.5 degrees of freedom, whose time-1 map is a twist map.

We only remark that how the factor  $\lambda_a^\sharp$  is determined. We first determine the  $\lambda_a$  before any reduction of order in Section 2.1. During each reduction of order, the value  $\lambda_a$  is modified (see Section 3.8) due to the refinement of the frequency and refined normal form. Eventually, after all the reduction of order, the system restricted on the NHIC is a system of two degrees of freedom and the frequency vector  $\omega_a$  is the rotation vector of the Aubry-Mather sets on the cylinder. Hence the  $\lambda_a^\sharp$  is naturally determined by the equation  $\alpha(\mathcal{L}_\beta(\omega_a^\sharp)) = E$ , where  $E$  is the fixed energy level.

**4.2. Induction around the strong double resonances.** In this section, we perform the reduction of order around a strong double resonance. A strong double resonance may appear during the  $j$ -th step of reduction of order. Without loss of generality and for simplicity of notations, we assume we encounter a strong double resonance point before any reduction of order.

The reduction scheme for strong double resonance is different from that in the single resonance regime since different Hamiltonian normal forms are used. To match two different regimes, we control the frequency vector  $\omega_a^j$  for each  $j$ , such that they have the same choice of rational numbers in the two different regimes. After totally  $n-2$  steps of reduction we get  $n-2$  linearly independent integer vectors  $k^1, \dots, k^{n-2} \in \mathbb{Z}^n \setminus \{0\}$  which determines a curve  $\omega_a^\sharp = \lambda_a^\sharp(a, \frac{p_2}{q_2}, \frac{p_3}{q_3}, \dots, \frac{p_n}{q_n}) \in H_1(\mathbb{T}^n, \mathbb{R})$  as  $a$  varies, where  $\lambda_a^\sharp$  is such that  $\alpha(\partial\beta(\omega_a^\sharp)) = E$  for some fixed energy level  $E$ . This frequency vector  $\omega_a^\sharp$  the rotation vectors of Aubry-Mather sets lying on a NHIC of dimension two (for the Poincaré return map that is twist) where the NHIC is known to exist. Since the same rotation vectors determine the same Aubry-Mather sets, in the overlapping region that can be handled as both single and double resonance, the two reduction schemes give the same result. For some  $a$ , there might be one more integer vector  $k''$  with norm bounded by a constant independent of any  $\delta$ 's that is perpendicular to  $\omega_a^\sharp$ .

The procedure is divided into the following steps: KAM normal form, linear symplectic transformation, shear transformation, the reduction of order and undoing the shear.

**4.2.1. The KAM normal form.** In Section 3.7, 3.10, we have shown how to perform the first and second steps of reduction of order. In the following, we describe the procedure inductively. For each  $j$ ,  $\omega_a^j$  has two integer vectors  $k''^j, k'^j \in \mathbb{Z}^{n-j}$  perpendicular to it. These integer vectors are transformed to vectors  $k'', k^1, \dots, k^{j+1}$  that are perpendicular to  $\omega_a^j$  through

$$(4.10) \quad k^t = M_0^t (\dots (M_{j-2}^t (\pi_{+2} (M_{j-1}^t \pi_{+2} (k^j)^t))))$$

with corresponding superscripts and  $k''^j$  gives rise to the same  $k'' = k''^1$  for all  $j$ .



In the following, we define  $j = j + 2 \in \{2, 3, \dots, n - 1\}$ . Here  $j$  counts the time of reduction of order and  $j$  counts the number of resonance relations. We perform KAM iterations to as we did in Proposition 2.2 and 3.2 to get the following normal form

$$(4.11) \quad \begin{aligned} H_j = & \frac{1}{\sqrt{\varepsilon}} \langle \omega_a^j, Y \rangle + \frac{1}{2} \langle AY, Y \rangle + P_I(Y) \\ & + \sum_{i=2}^j \delta_i V_i(\langle k'', x \rangle, \langle k'^1, x \rangle, \langle k'^2, x \rangle, \dots, \langle k'^{i-1}, x \rangle) \\ & + \delta_{j+1} R_{j,I}(x) + \varepsilon^\sigma R_{j,II}(x, Y) \end{aligned}$$

where  $\delta_{j+1}$  can be chosen as small as we wish, and  $\delta_2 \gg \delta_3 \gg \dots \gg \delta_j$  are chosen and fixed as our inductive hypothesis.

*Our goal in this section is to determine an upper bound for  $\delta_{j+1}$  and will fix  $\delta_{j+1}$  provided the upper bound is satisfied in order to proceed to determining  $\delta_{j+2}$ .*

**4.2.2. The linear symplectic transformation.** We construct a matrix  $M_j \in SL(n, \mathbb{Z})$  whose first  $j$  rows are exactly  $k'', k'^1, k'^2, \dots, k'^{j+1}$ . Readers can skip the next paragraph if s/he is familiar with this fact.

We denote by  $M_2 = M''M'$  and  $M_3 = \mathbf{M}_2''\mathbf{M}_2'\mathbf{M}_1'$ ,  $M_j = \mathbf{M}_{j+1}''(\mathbf{M}_{j+1}'\mathbf{M}_j' \cdots \mathbf{M}_1')$  where  $\mathbf{M}_i'$  is the enlargement of  $M_i'$  into a matrix in  $SL(n, \mathbb{Z})$ ,  $\mathbf{M}_i' = (\pi_{+2})^{i-1}M_i'$ ,  $i = 2, \dots, j$  and  $\mathbf{M}_{j+1}''$  has the following form

$$\mathbf{M}_{j+1}'' = \begin{bmatrix} & \tilde{\mathbf{k}}''^{j+1} & \\ 0_{(j+1) \times 1} & \text{id}_{j+1} & 0_{(j+1) \times (n-j-2)} \\ *_{(n-j-2) \times 1} & 0_{(n-j-2) \times 2} & *_{(n-j-2)^2} \end{bmatrix}$$

where  $\tilde{\mathbf{k}}''^{j+1}$  is defined from the relation  $\tilde{\mathbf{k}}''^{j+1}\mathbf{M}_{j+1}' \cdots \mathbf{M}_2'M' = k''$  and the  $*$  entries are defined as follows. In  $\mathbf{M}_{j+1}''$ , we remove the second  $\dots (j+1)$ -th columns and rows to get a matrix of  $(n-j-1) \times (n-j-1)$ . Then from the resulting first row, we find  $n-j-2$  integer vectors spanning unit volume with the first row. The matrix  $\mathbf{M}_2''$  determined in this way is in  $SL(n, \mathbb{Z})$ . We know by construction that the first  $j$  rows of the matrix  $M_j$  are  $k'', k'^1, k'^2, \dots, k'^{j+1}$  respectively. We relabel these vectors by  $\mathbf{k}^1, \dots, \mathbf{k}^j$  respectively and denote by  $\mathbf{K}_j$  the matrix in  $\mathbb{Z}^{j \times n}$  formed by these vectors as rows.

The matrix  $M_j$  induces a symplectic transformation

$$\mathfrak{M}_j : (x, y) \mapsto (M_j x, M_j^{-t} y).$$

We denote  $A_j = M_j A M_j^t$  and the frequency vector  $\omega_a^j$  is transformed to  $M_j \omega_a^j = (0_j, \hat{\omega})$  for some non resonant vector  $\hat{\omega} \in \mathbb{R}^{n-j-2}$ . Then the  $(i, j)$ -th entry of  $A_j$  is given by  $a_{ij} = \mathbf{k}^i A(\mathbf{k}^j)^t$ . The Hamiltonian (4.11) under the transformation becomes

$$(4.12) \quad \begin{aligned} \mathfrak{M}_j^{-1*} H_j = & \frac{1}{\sqrt{\varepsilon}} \langle \hat{\omega}_{n-j}, \hat{Y}_{n-j} \rangle + \frac{1}{2} \langle A_j Y, Y \rangle + \sum_{i=2}^j \delta_i V_i(x_1, x_2, \dots, x_i) \\ & + P_I(Y) + \delta_{j+1} R_{j,I}(x) + \varepsilon^\sigma R_{j,II}(x, Y). \end{aligned}$$

where  $P_I(Y) = \mathfrak{M}_j^{-1} P_1(Y)$ , similarly for  $R_{j,I}$  and  $R_{j,II}$ . The genericity condition that we need is the following which can be proved generic following the same argument as **(H2)**<sub>2</sub>.

**(H2)<sub>j+1</sub>**: Assume **(H2)<sub>j</sub>**. For generic  $V_{j+2} \in C^r(\mathbb{T}^3, \mathbb{R})$  ( $r > 5$ ) and sufficiently small  $\delta_{j+2}$ , the action minimizing invariant measures with rotation vector  $\nu(1, 0, \dots, 0)$   $|\nu| \in [\gamma, \infty)$ ,  $\forall \gamma > 0$  of the system  $\frac{1}{2}\langle A_j Y, Y \rangle + \sum_{i=2}^j \delta_i V_i(x_1, x_2, \dots, x_i)$  are supported on one hyperbolic periodic orbits except for finitely many  $\nu$ 's for which there are two minimal hyperbolic periodic orbits. These periodic orbits make up finitely many pieces NHICs.

**4.2.3. The shear transformation.** We introduce a shear transformation in this section to block diagonalize the quadratic form  $\langle A_j Y, Y \rangle$ . Let

$$A_j = \begin{bmatrix} \tilde{A}_j & \check{A}_j \\ \check{A}_j^t & \hat{A}_j \end{bmatrix}, \quad S_j = \begin{bmatrix} \text{id}_j & 0 \\ -\check{A}_{I,j}^t \tilde{A}_j^{-1} & \text{id}_{n-j} \end{bmatrix},$$

$B_j = \hat{A}_j - \check{A}_j^t \tilde{A}_j^{-1} \check{A}_j$  where  $\tilde{A}_j \in \mathbb{R}^{j^2}$ ,  $\check{A}_j \in \mathbb{R}^{j \times (n-j)}$  and  $\hat{A}_j \in \mathbb{R}^{(n-j)^2}$ . It can be verified that  $S_{I,j} A_j S_{I,j}^t = \text{diag}(\tilde{A}_j, B_j)$ . We introduce a linear transformation

$$\mathfrak{S}_{I,j} : (x, Y) \mapsto (S_{I,j} x, S_{I,j}^{-t} Y) := (x, y),$$

which transforms the Hamiltonian into the following defined on  $T^*\mathbb{T}^j \times \mathbb{R}^{2(n-j)}$

$$(4.13) \quad \begin{aligned} \mathfrak{S}_{I,j}^{-1*} \mathfrak{M}_j^{-1*} H_j &= \left[ \frac{1}{2} \langle \tilde{A}_j \tilde{y}_j, \tilde{y}_j \rangle + \sum_{i=2}^j \delta_i V_i(x_1, x_2, \dots, x_i) \right] \\ &+ \frac{1}{\sqrt{\varepsilon}} \langle \hat{\omega}_{n-j}, \hat{y}_{n-j} \rangle + \frac{1}{2} \langle B_j \hat{y}_{n-j}, \hat{y}_{n-j} \rangle \\ &+ P_I(S_{I,j}^t y) + \mathfrak{S}_{I,j}^{-1*} (\delta_{j+1} R_{j,I}(x) + \varepsilon^\sigma R_{j,II}(x, y)), \end{aligned}$$

where  $\tilde{y}_j = (y_1, \dots, y_j)$  and  $\hat{y}_{n-j} = (y_{j+1}, \dots, y_n)$ . The norms of  $B_j, R_{j,I}, R_{j,II}$  depend only on  $\delta_2, \dots, \delta_j$ , but not on  $\delta_{j+1}$ .

Next, we further block diagonalize the matrix  $\tilde{A}_j$ . We decompose  $\tilde{A}_j$  into the form

$$\tilde{A}_j = \begin{bmatrix} \tilde{A} & \mathbf{a}_3^t & & \\ \mathbf{a}_3 & a_{33} & \mathbf{a}_4^t & \vdots \\ & \mathbf{a}_4 & a_{44} & \\ \dots & & & \ddots & \mathbf{a}_j^t \\ & & & \mathbf{a}_j & a_{jj} \end{bmatrix}$$

where  $\tilde{A}$  is the first  $2 \times 2$  block,  $a_{ii}$  are the diagonal entries and  $\mathbf{a}_i \in \mathbb{R}^{i-1}$  is a vector formed by the  $i-1$  entries in the  $i$ -th row to the left of the diagonal. We next form matrices

$$(4.14) \quad \begin{aligned} T_i &= \begin{bmatrix} \text{id}_i & 0 & 0 \\ -\mathbf{a}_{i+1} \tilde{A}_i^{-1} & 1 & 0 \\ 0 & 0 & \text{id}_{n-i-1} \end{bmatrix}, \\ S_{II,j} = T_2 \cdots T_{j-1} &= \begin{bmatrix} \text{id}_2 & 0 & 0 & 0 & 0 \\ -\mathbf{a}_3 \tilde{A}_2^{-1} & 1 & 0 & 0 & 0 \\ & \dots & \ddots & 0 & 0 \\ & & -\mathbf{a}_j \tilde{A}_{j-1}^{-1} & 1 & 0 \\ 0 & 0 & 0 & 0 & \text{id}_{n-j} \end{bmatrix}. \end{aligned}$$

We denote

$$b_i = a_{ii} - \mathbf{a}_i \tilde{A}_{i-1}^{-1} \mathbf{a}_i^t = \mathbf{k}^i (A - (A \mathbf{K}_{i-1}^t) (\mathbf{K}_{i-1} A \mathbf{K}_{i-1}^t)^{-1} (\mathbf{K}_{i-1}^t A)) (\mathbf{k}^i)^t$$

where  $\mathbf{K}_{i-1}$  is a matrix in  $\mathbb{Z}^{(i-1) \times n}$  whose rows are  $\mathbf{k}^1, \dots, \mathbf{k}^{i-1}$ . We denote the first entry of the vector  $\mathbf{a}_i \tilde{A}_{i-1}^{-1}$  by  $s_i$  so the first column of  $S_{II,j}$  is  $(1, 0, -s_3, \dots, -s_j, \hat{0}_{n-j})$ .

The same Gauss least square argument as in Section 3.10.6 gives us

$$(4.15) \quad b_i = \text{const}_{b,i} |\mathbf{k}^i|^2 = \text{const}_{b,i} |\mathbf{k}^{i-1}|^2, \quad s_i = \text{const}_{s,i} |\mathbf{k}^i| = \text{const}_{s,i} |\mathbf{k}^{i-1}|,$$

where  $\text{const}_{b,i} (> 0)$  and  $\text{const}_{s,i}$  do not depend on  $\delta_i$  (but might depend on  $\delta_k$ ,  $k < i$ ).

With these definitions, we diagonalize the matrix  $A$ . Let  $S_j = S_{II,j} S_{I,j}$  then

$$S_j A_j S_j^t = \text{diag}\{\tilde{A}, b_3, \dots, b_j, B_j\}.$$

We introduce the transformation  $\mathfrak{S}_j: (x, y) \rightarrow (S_j x, S_j^{-t} y)$  and obtain a Hamiltonian defined on  $T^* \mathbb{T}^2 \times \mathbb{R}^{2(n-2)}$

$$(4.16) \quad \begin{aligned} \mathfrak{S}_j^{-1*} \mathfrak{M}_j^{-1*} H_j = & \frac{1}{2} \langle \tilde{A} \tilde{y}, \tilde{y} \rangle + \delta_2 V_2(\tilde{x}) + \sum_{i=3}^j \left( \delta_i V_i \left( S_{II,j}^{-1}(\tilde{x}_i, \hat{0}_{n-i}) \right) + \frac{b_i}{2} y_i^2 \right) \\ & + \frac{1}{\sqrt{\varepsilon}} \langle \hat{\omega}_{n-j}, \hat{y}_{n-j} \rangle + \frac{1}{2} \langle B_j \hat{y}_{n-j}, \hat{y}_{n-j} \rangle \\ & + \mathfrak{S}_j^{-1*} (P_I(y) + \delta_{j+1} R_{j,I}(x) + \varepsilon^\sigma R_{j,II}(x, y)). \end{aligned}$$

where  $\tilde{x}_i = (x_1, \dots, x_i)$  and the *hat* notation is standard.

**4.2.4. The reduction of order and undo-shear transformation.** We perform the order reduction to the system (4.16). The general idea is always to separate the first two degrees of freedom from the system and apply Theorem 3.4 to find a NHIC in the subsystem with homology class  $g = (1, 0) \in H_1(\mathbb{T}^2, \mathbb{Z})$ . Restricted to the NHIC by applying Theorem 3.2, we reduce the number of degrees of freedom by one. We fix the homology class  $g = (1, 0) \in H_1(\mathbb{T}^2, \mathbb{Z})$  since we have zero as the second component of the frequency vector in (4.5) after applying  $M'_j$ . The first two steps of the reduction is done in Section 3.7 and 3.10. Since the shear transformation  $S_j \notin SL(n, \mathbb{Z})$  and it changes the resonance relations, we always need an undo-shear transformation after one reduction to get a subsystem of two degrees of freedom defined on  $T^* \mathbb{T}^2$  rather than  $T^* \mathbb{T}^1 \times \mathbb{R}^2$ . Suppose that after the  $i$ -th reduction,  $i = 1, 2, \dots, j-1$ , we obtain a Hamiltonian defined on  $T^* \mathbb{T}^1 \times \mathbb{R}^{2(n-i-2)}$

$$(4.17) \quad \begin{aligned} \bar{H}_{j,i} = & \left[ \tilde{h}_{i+1}(I_{i+1}) + \delta_{i+2} Z_{i,i+2}(I_{i+1}, \varphi_{i+1}, x_{i+2}) + \frac{1}{2} b_{i+2} y_{i+2}^2 \right] \\ & + \sum_{k=i+3}^j \left( \frac{b_k}{2} y_k^2 + \delta_k Z_{i,k}(I_{i+1}, \varphi_{i+1}, x_{i+2}, \dots, x_k) \right) \\ & + \frac{1}{\sqrt{\varepsilon}} \langle \hat{\omega}_{n-j}, \hat{y}_{n-j} \rangle + \frac{1}{2} \langle B_j \hat{y}_{n-j}, \hat{y}_{n-j} \rangle + (\bar{P}_{I,i} + \delta_{j+1} \bar{R}_{i,I} + \varepsilon^\sigma \bar{R}_{i,II}) \end{aligned}$$

where  $Z_{i,k}, \bar{P}_{I,i}, \bar{R}_{i,I}, \bar{R}_{i,II}$  are the  $i$ -th reduction from  $V_k, P_I, R_I, R_{II}$  respectively.

Now the bracketed subsystem is only defined on  $T^* \mathbb{T}^1 \times \mathbb{R}^2$  but not on  $T^* \mathbb{T}^2$ . The new undo-shear matrix

$$\underline{S}_i = \begin{bmatrix} 1 & 0 & 0 \\ s_{i+2} & 1 & 0 \\ 0 & 0 & \text{id}_{n-i-2} \end{bmatrix}$$

transform the bracketed term above to

$$\tilde{h}_{i+1}(J_{i+1} + s_{i+2} y_{i+2}) + \frac{1}{2} b_{i+2} y_{i+2}^2 + \delta_{i+2} Z_{i,i+2}(J_{i+1} + s_{i+2} y_{i+2}, \varphi_{i+1}, x_{i+2} - s_{i+2} \varphi_{i+1}),$$

where  $J_{i+1} = I_{i+1} - s_{i+2}y_{i+1}$  and  $x_{i+2} = s_{i+2}\varphi_{i+1} + x_{i+2}$ . Now we obtain a system of two degrees of freedom defined on  $T^*\mathbb{T}^2$  and we can perform the reduction to get a system of one degree of freedom denoted by  $\tilde{h}_{i+2}(I_{i+2})$ , and a system (4.17) with  $i$  updated to  $i + 1$ . We repeat this procedure until  $i = j - 1$ . The NHIC obtained by applying Theorem 3.4 to the  $j - 1$ -th step determines an upper bound for  $\delta_{j+1}$  in order for Theorem 3.1, 3.2 to be applicable. This completes the induction for  $j$ .

**4.3. Reduction around complete resonance.** Suppose we have completed all the reduction of orders hence determined all the  $\delta$ 's. We set  $j = n - 1$  in the previous section. In this section, we list some final normal forms that will be used for construction of diffusing orbit in the later sections. As we did in Section 3.5, we distinguish low energy ( $|Y|$  bounded by constants independent of  $\varepsilon$  but allowed to depend on  $\delta$ 's) and high energy region and restrict our attention to the low energy region. So we can absorb the  $P_I$  term into the perturbation  $\varepsilon^\sigma R_{II}$  using (2.16).

Removing the last row and column of  $A_{n-1} = M_{n-1}AM_{n-1}^t$  we get a matrix  $A \in GL(n-1, \mathbb{R})$ . The  $(i, j)$ -th entry of  $A$  is  $a_{ij} = \mathbf{k}^i \mathbf{A}(\mathbf{k}^j)^t$ . With these notations,  $(n-2)$  steps of KAM iterations lead to (4.11) with  $j = n-3$  and  $j = j+2 = n-1$ . Followed by a linear symplectic transformation  $\mathfrak{M}_{n-1} : (x, Y) \mapsto (M_{n-1}x, M_{n-1}^{-t}Y)$  one obtains

$$(4.18) \quad H_{n-1} = \frac{1}{\sqrt{\varepsilon}} \langle \omega_n, Y_n \rangle + \frac{1}{2} \langle A_{n-1}Y, Y \rangle + \sum_{i=2}^{n-1} \delta_i V_i(x_1, x_2, \dots, x_i) + \delta_n R(x).$$

Slightly different from the previous procedure, here we first perform a standard energetic reduction to reduce it to a system of  $n - 1/2$  degrees of freedom. As  $\omega_n \neq 0$  and  $\varepsilon > 0$  is very small, one has the function  $Y_n(x, x_n, y)$  as the solution of the equation

$$H_{n-1}(x, x_n, y, Y_n(x, x_n, y)) = E^* > \min \alpha_{H_{n-1}},$$

which takes the form  $Y_n = -Y_\delta \frac{\sqrt{\varepsilon}}{\omega_n}$ , where

$$(4.19) \quad Y_\delta = \frac{1}{2} \langle Ay, y \rangle + \sum_{j=2}^{n-1} \delta_j V_j(x_1, \dots, x_j) + \delta_n R(x, x_n, y)$$

where we update the notation  $x = (x_1, \dots, x_{n-1})$ ,  $y = (Y_1, \dots, Y_{n-1})$ .

We next apply  $\mathfrak{S}_{II,2}$  in (4.14) with  $j = n - 1$  to block diagonalize the quadratic form  $A$  where we remove the last row and column of  $S_{II,2}$ . The Hamiltonian system has the normal form defined on  $T^*\mathbb{T}^2 \times \mathbb{R}^{2(n-3)} \times \mathbb{T}^1$

$$(4.20) \quad \begin{aligned} Y_\delta := \mathfrak{S}_{II,2}^{-1*} Y_\delta &= \frac{1}{2} \langle \tilde{A} \tilde{y}, \tilde{y} \rangle + \delta_2 V_2(\tilde{x}) \\ &+ \sum_{i=3}^{n-1} \left( \frac{b_i}{2} y_i^2 + \delta_i V_i \left( S_{II,j}^{-1}(\tilde{x}_i, \hat{0}_{n-i-1}) \right) \right) + \delta_n R, \end{aligned}$$

where  $\tilde{x} = (x_1, x_2)$ ,  $\tilde{x}_i = (x_1, \dots, x_i)$ . Similarly for  $\tilde{y}, \tilde{y}_i$ . With this Hamiltonian, we perform the reduction of order as we did in the previous section. During the  $i$ -th step,  $i = 1, 2, \dots, n-2$ , we have the following normal form defined on  $T^*\mathbb{T}^2 \times \mathbb{R}^{2(n-i-2)}$

after an undo-shear transform

$$\begin{aligned}
 \bar{Y}_{i,\delta} = & \left[ \tilde{h}_{i+1}(J_{i+1} + s_{i+2}y_{i+2}) + \frac{1}{2}b_{i+2}y_{i+2}^2 \right. \\
 (4.21) \quad & \left. + \delta_{i+2}Z_{i,i+2}(J_{i+1} + s_{i+2}y_{i+2}, \varphi_{i+1}, x_{i+2} - s_{i+2}\varphi_{i+1}) \right] \\
 & + \sum_{k=i+3}^{n-1} \left( \frac{b_k}{2}y_k^2 + \delta_k Z_{i,k}(J_{i+1}, \varphi_{i+1}, x_{i+2}, x_{i+3}, \dots, x_k) \right) + \delta_n \bar{R}_i.
 \end{aligned}$$

**Remark 4.1** (Smoothness requirement). We need to perform  $n - 2$  steps of order reduction. One step of reduction (either for single or for double resonance) makes the reduced system lose two derivatives. In each KAM normal form, the  $R_I$  term is obtained from  $P(x, y)$  by taking the high frequency Fourier modes without taking any derivatives and the  $R_{II}$  loses two derivatives. The genericity conditions  $(\mathbf{H1})_j, (\mathbf{H2})_j$  have nothing to do with the  $R_{II}$  terms hence the  $C^5$  smoothness are always guaranteed for  $C^r$ -Hamiltonian with  $r \geq 2n$ . The requirement of smoothness here coincides with the threshold of smoothness so that KAM tori exist. The KAM theorem holds when perturbation is  $C^{2n+\delta}$ -small and recently, it was shown in [C11] that KAM theory does not apply if the perturbation is small only in  $C^{2n-\delta}$ -topology.

**4.4. Admissible frequency path.** In this section, we provide an algorithm to move the second, third ... components of the frequency vector.

**Definition 4.1** (admissible frequency segment). A frequency segment of the form  $\omega_a^\sharp = \lambda_a^\sharp(a, \frac{p_2}{q_2}, \frac{p_3}{q_3}, \dots, \frac{p_n}{q_n})$  is said admissible (up to permutation of entries) for the Hamiltonian system  $H$  defined in (1.1) which satisfies the generic conditions  $(\mathbf{H1})_j, (\mathbf{H2})_j$  along the segment, if there exists a hierarchy of numbers

$$\mu_0 \gg \mu_1 \gg \dots \gg \mu_{n-3}$$

and a Diophantine vector  $\hat{\omega}_{n-3}^* \in \text{DC}(n-2, \alpha, \tau)$  such that

- (1)  $\left| \frac{p_{j+3}}{q_{j+3}} - \omega_{j+3}^* \right| < \mu_j, \quad j = 0, 1, \dots, n-3,$
- (2)  $\mu_{j+1}$  is determined by  $\delta_{j+1}$  through Lemma 2.3 with  $K = \delta_{j+1}^{-1/2}$ , and  $\delta_{j+1} < \Delta(\mathbf{k}'^j)$  is chosen arbitrarily, where  $\Delta(\mathbf{k}'^j)$  is given by Lemma 3.3 and  $\mathbf{k}'^j$  is determined by  $\frac{p_2}{q_2}, \dots, \frac{p_{j+3}}{q_{j+3}}$  using (4.2) inductively.

**Definition 4.2** ( $\rho$ -accessibility). For any given  $\rho$ , we say two vectors  $\omega^i$  and  $\omega^f$  are  $\rho$ -accessible to each other if a  $\rho$ -neighborhood of  $\omega^i$  can be joined to that of  $\omega^f$  by piecewise admissible frequency segments.

We will show that the accessibility implies the existence of diffusing orbits. In this section we prove the following theorem

**Theorem 4.1.** Given any two vectors  $\omega^i, \omega^f \in \mathbb{R}^n$  and any small number  $\rho > 0$ , there are two vectors  $\omega^i, \omega^f \in \mathbb{R}^n$  that are accessible to each other and satisfying

$$|\omega^i - \omega^i| < \rho, \quad |\omega^f - \omega^f| < \rho.$$

To prove this theorem we need the following number theoretical lemma.

**Lemma 4.1.** Given any  $\rho > 0, \tau > n$  and any two frequency vectors  $\omega^i \neq \omega^f \in \partial h^{-1}(h^{-1}(E))$ , there exist constant  $\alpha > 0$  and two vectors

$$\omega^{*i} = (\omega_1^{*i}, \dots, \omega_n^{*i}), \quad \omega^{*f} = (\omega_1^{*f}, \dots, \omega_n^{*f})$$

satisfying the following

$$|\omega^i - \omega^{*i}| < \rho, \quad |\omega^f - \omega^{*f}| < \rho$$

and the  $(n+1)$  vectors  $(\omega_1^{*f}, \dots, \omega_j^{*f}, \omega_{j+1}^{*i}, \dots, \omega_n^{*i}) \in \text{DC}(n+1, \alpha, \tau)$  for all  $j = 0, 1, 2, \dots, n$ .

*Proof.* Fix  $\rho > 0, \tau > n$ . We prove the lemma by induction from  $j+1$  to  $j$ . First, for  $j = n$ , it is easy to find two Diophantine numbers  $\omega_n^i, \omega_n^f$ . Suppose we already have that

$$(\omega_{j+1}^{*i}, \dots, \omega_n^{*i}) \in \text{DC}(n-j+1, \alpha, \tau).$$

We claim that there are numbers  $\omega_j^{*i}$  and  $\omega_j^{*f}$  satisfying  $|\omega_j^{*i} - \omega_j^i| < \rho, |\omega_j^{*f} - \omega_j^f| < \rho$ , and

$$(\omega_j^{*i,f}, \omega_{j+1}^{*i}, \dots, \omega_n^{*i}) \in \text{DC}(n-j+2, \alpha, \tau)$$

for sufficiently small  $\alpha > 0$ . Indeed, by assumption we already have

$$\left| \langle (1, \hat{\omega}_{n-j}^{*i}), \mathbf{k}_{n-j+1} \rangle \right| \geq \frac{\alpha}{|\hat{\mathbf{k}}_{n-j}|^\tau}, \quad \forall \mathbf{k}_{n-j+1} \in \mathbb{Z}^{n-j+1} \setminus \{0\}.$$

We want to show that all those  $\omega_j \in \mathbb{R}$  which satisfy the condition

$$(4.22) \quad |\langle (1, \omega_j, \omega_{n-j}^{*i}), \mathbf{k}_{n-j+2} \rangle| \geq \frac{\alpha}{|\hat{\mathbf{k}}_{n-j+1}|^\tau}, \quad \forall (k_j, \mathbf{k}_{n-j+2}) \in \mathbb{Z}^{n-j+2} \setminus \{0\}$$

form a  $\rho$ -dense set provided  $\alpha$  is small enough. Given  $\hat{\mathbf{k}}_{n-j+1}$ , we consider all  $k_j$  and  $\omega_j^\dagger$  satisfying

$$k_j \omega_j^\dagger + \langle (1, \hat{\omega}_{n-j}^{*i}), \hat{\mathbf{k}}_{n-j+1} \rangle = 0.$$

Formula (4.22) is satisfied automatically when  $k_j = 0$  so we assume  $k_j \neq 0$ . In order to guarantee (4.22) we need to remove an interval of measure  $\frac{2\alpha}{k_j(|\hat{\mathbf{k}}_{n-j}| + |k_j|)^\tau}$  centered at  $\omega_j^\dagger$  so that (4.22) is satisfied for all  $\omega_j$  in the complement for this  $k_j$ . The total measure of these intervals removed when  $k_j$  ranges over  $\mathbb{Z}$  is

$$\sum_{k_j} \frac{2\alpha}{|k_j|(|\hat{\mathbf{k}}_{n-j}| + |k_j|)^\tau} \leq 2 \int_1^\infty \frac{2\alpha}{x(|\hat{\mathbf{k}}_{n-j}| + x)^\tau} dx.$$

Next the total measure of these intervals when  $\hat{\mathbf{k}}_{n-j}$  ranges over  $\mathbb{Z}^{n-k}$  is

$$\begin{aligned} \sum_{\mathbf{k}_{n-j}} \sum_{k_j} \frac{4\alpha}{|k_j|(|\hat{\mathbf{k}}_{n-j}| + |k_j|)^\tau} &\leq \sum_{\mathbf{k}_{n-j}} \int_1^\infty \frac{4\alpha}{x(|\hat{\mathbf{k}}_{n-j}| + x)^\tau} dx \\ &\leq \int_{\mathbb{S}^{n-j-1}} \int_1^\infty \int_1^\infty \frac{4\alpha}{x(r+x)^\tau} dx r^{n-j-1} dr d\mathbb{S}^{n-j-1} \\ &\stackrel{y=x/r}{=} 4\alpha C \int_1^\infty r^{n-j-\tau-1} \int_{1/r}^\infty \frac{1}{y(1+y)^\tau} dy dr \end{aligned}$$

where the constant  $C = \frac{2\pi^{(n-j-1)/2}}{\Gamma((n-j-1)/2)}$  is the area of the sphere  $\mathbb{S}^{n-j-1}$ . The inner integral converges for large  $y$  and has the asymptote  $\log r$  for  $r$  large and  $y$  close to  $1/r$ . Hence the iterated integral can be estimated as

$$\int_1^\infty r^{n-j-\tau-1} \int_{1/r}^\infty \frac{1}{y(1+y)^\tau} dy dr \leq 2 \int_1^\infty r^{n-j-\tau-1} (\log r + \text{const}) dr$$

where the right-hand-side is convergent since  $\tau > n$ . The assertion above is proven if  $\alpha > 0$  is chosen small enough.  $\square$

In virtue of this lemma, we are able to prove Theorem 4.1.

*Proof of Theorem 4.1.* As our strategy, we start with an admissible frequency segment of the form  $\omega_a = (a, \frac{p_2}{q_2}, \frac{p_3}{q_3}, \dots, \frac{p_n}{q_n})$  with numbers  $\mu_0 \gg \mu_1 \gg \dots \gg \mu_n$  such that

$$\left| \omega_{j+3}^{*i} - \frac{p_{j+3}}{q_{j+3}} \right| < \mu_j, \quad j = 0, 1, \dots, n-3.$$

Next, we move the number  $a$  to the point  $\frac{p_{n+1}}{q_{n+1}}$  satisfying  $|\frac{p_{n+1}}{q_{n+1}} - \omega_1^{*i}| < \mu_{n-2}$  for some number  $\mu_{n-2}$ . We want the line segment  $\omega_a = (a_2, \frac{p_3}{q_3}, \dots, \frac{p_n}{q_n}, \frac{p_{n+1}}{q_{n+1}})$  to be admissible such that we can move  $a_2$  to the point  $\frac{p_{n+2}}{q_{n+2}}$  satisfying  $|\frac{p_{n+2}}{q_{n+2}} - \omega_2^{*i}| < \mu_{n-1}$  for some number  $\mu_{n-1}$ , and so on.

We need to show how to determine the rational numbers  $\frac{p_{j+3}}{q_{j+3}}$  as well as the numbers  $\mu_j$  for  $j = 0, 1, \dots, 2n-3$  inductively. We claim:

*for given  $\omega^{*i}$  and  $\omega^{*f}$ , there are rational numbers  $\frac{p_2}{q_2}, \frac{p_3}{q_3}, \dots, \frac{p_{2n}}{q_{2n}}$  as well as numbers  $\mu_0 \gg \mu_1 \gg \dots \gg \mu_{2n-3}$  such that*

$$(4.23) \quad \begin{aligned} \left| \omega_{j+3}^{*i} - \frac{p_{j+3}}{q_{j+3}} \right| &< \mu_j, \quad j = 0, 1, \dots, n-3 \\ \left| \omega_{j+3}^{*f} - \frac{p_{j+3}}{q_{j+3}} \right| &< \mu_j, \quad j = n-4, 1, \dots, 2n-3. \end{aligned}$$

Moreover, the  $n-1$  vectors

$$\omega_j := \left( a_{j+1}, \frac{p_{j+2}}{q_{j+2}}, \frac{p_{j+3}}{q_{j+3}}, \dots, \frac{p_{n+j}}{q_{n+j}} \right), \quad j = 0, 1, \dots, n-3$$

are all admissible.

This implies the accessibility of the two vectors  $\omega^{*i}$  and  $\omega^{*f}$ .

Let us first show how to determine the numbers  $\mu_0, \mu_1$ , then we explain how to determine all the remaining numbers inductively.

For the frequency segment  $\omega_1 = (a_1, \frac{p_2}{q_2}, \frac{p_3}{q_3}, \dots, \frac{p_n}{q_n})$  to be admissible, we first choose rational numbers  $\frac{p_2}{q_2}, \frac{p_3}{q_3}$  that are  $\rho$  close to  $\omega_2^{*i}$  and  $\omega_3^{*i}$  respectively. Next we get  $\mu_0$  determined by  $\delta_0 < \Delta(\mathbf{k}'_0)$  (where  $\mathbf{k}'_0$  is determined by  $\frac{p_2}{q_2}, \frac{p_3}{q_3}$ ), which is in turn determined by item (2) of Proposition 4.1 (see also item (2) of Definition 4.1). Next  $\frac{p_4}{q_4}$  can be chosen to be any rational number satisfying  $|\frac{p_4}{q_4} - \omega_4^{*i}| < \mu_1$  according to item (1) of Definition 4.1.

Next, applying the same procedure we determine a number  $\mu_{1,2}$  by the integer vector  $\mathbf{k}'_{1,1}$  which is determined by  $\frac{p_4}{q_4}, \frac{p_3}{q_3}, \frac{p_2}{q_2}$  using (4.2). We can choose  $\frac{p_5}{q_5}$  to be any rational number satisfying  $|\frac{p_5}{q_5} - \omega_5^{*i}| < \mu_{1,2}$ . Here the first subscript 1 of  $\mu_{1,2}, \mathbf{k}'_{1,1}$  is inherited from  $a_1$  meaning the component that is moving and the second subscript is the usual subscript for the  $\mu$  counting the number of reduction of order.

However, we need to guarantee that the second frequency segment  $\omega_2 = (a_2, \frac{p_3}{q_3}, \frac{p_4}{q_4}, \dots, \frac{p_{n+1}}{q_{n+1}})$  to be admissible. For  $\omega_2$ , we determine a number  $\mu_{2,2}$  by the integer vector  $\mathbf{k}'_{2,1}$  which is determined by  $\frac{p_4}{q_4}, \frac{p_3}{q_3}$  using (4.2). We can choose  $\frac{p_5}{q_5}$  to be any rational number satisfying  $|\frac{p_5}{q_5} - \omega_5^{*i}| < \mu_{2,2}$ .

To satisfy the above two restrictions simultaneously, we denote  $\mu_2 = \min\{\mu_{1,2}, \mu_{2,2}\}$  and choose  $\frac{p_5}{q_5}$  to be a rational number satisfying  $|\frac{p_5}{q_5} - \omega_5^{*i}| < \mu_2$ . Once  $\frac{p_5}{q_5}$  is chosen, it is fixed.

Inductively, we determine  $\mu_{i,j}$  for  $\omega_i$ ,  $i = 1, 2, \dots, j$ ,  $j = 1, 2, \dots, 2n-3$ , from item (2) of Proposition 4.1 once  $\frac{p_{j+2}}{q_{j+2}}$  is given. We define  $\mu_j := \min_{i \leq j} \{\mu_{i,j}\}$  and choose  $\frac{p_{j+3}}{q_{j+3}}$  to be a rational number satisfying (4.23). This completes the proof.  $\square$

## 5. MAÑÉ SETS ALONG $s$ -RESONANCE AND AROUND $m$ -STRONG RESONANCE

To construct diffusion orbits, we are going to use variational method closely related to the Mather theory. Towards this goal, we need to know sufficient information on the structure of relevant Mañé sets.

Recall the procedure of reduction, there are two types of reduction, one is for single resonance, another one is for strong double resonance. Except for small neighborhoods of finitely many resonant points, along remaining intervals all  $n-3$  steps of reduction are the type of single resonant case. We call it  $s$ -resonance (always in single resonance). At those points we are encountered strong double resonance for the first time at  $m$ -th step of reduction, for some  $m \in \{1, \dots, n-2\}$ . We call this case  $m$ -strong resonance, as all follow-up reductions are the type of double resonant case.

**5.1. Mañé set along  $s$ -resonance.** As the first step, let us study the normal form (2.18) at single resonance to see where are the Mather sets we are concerned about. As analyzed in Section 3.2, the system (2.18) generically admits a NHIC  $\Pi_{n-1}^\delta$  which is diffeomorphic to  $\mathbb{T}^{n-1} \times \mathbb{R}^{n-1}$ . For the Lagrangian  $L_{H'}$  determined by  $H'$ , we claim that, for each  $c \in \mathcal{L}_{\beta_{H'}}(\omega_1, 0, \hat{\omega}_{n-2})$ , the Mather set stays on  $\mathcal{L}_{H'}(\Pi_{n-1}^\delta)$ . Indeed, for  $\delta = 0$ , the Mather set for such  $c$  is obviously located on  $\Pi_{n-1}^0$ , which is normally hyperbolic. For  $c \in \text{int } \mathcal{L}_{\beta_{H'}}(\omega_1, 0, \hat{\omega}_{n-2})$ , the Mañé set is the same as the Mather set. Because of upper-semi continuity, for small  $\delta$ , the Mañé set is in a small neighborhood of the manifold. On the other hand, due to the hyperbolic structure, any orbit cannot stay in a small neighborhood of  $\Pi_{n-1}^\delta$  unless this orbit entirely stays in  $\Pi_{n-1}^\delta$ . This verifies our claim. Restricted on this manifold, the Hamiltonian has  $(n-1)$ -degrees of freedom.

Next, let us study the normal form (2.19) at strong double resonance. Under the linear transformations  $\mathfrak{M}'$  and  $\mathfrak{M}''$

$$\begin{aligned} H'' = (\mathfrak{M}''^{-1} \mathfrak{M}'^{-1} \phi)^* H &= \frac{1}{2} \langle \tilde{A}(\tilde{Y} + \tilde{A}^{-1} \check{A} \hat{Y}_{n-2}), (\tilde{Y} + \tilde{A}^{-1} \check{A} \hat{Y}_{n-2}) \rangle + V(\tilde{x}) \\ &+ \frac{1}{2} \langle \hat{Y}_{n-2}, (\hat{A} - \check{A}^t \tilde{A}^{-1} \check{A}) \hat{Y}_{n-2} \rangle + \frac{1}{\sqrt{\varepsilon}} \langle \hat{\omega}_{n-2}, \hat{Y}_{n-2} \rangle \\ &+ \delta R(x, Y). \end{aligned}$$

Let  $L_{H''}$  be the Lagrangian determined by  $H''$  and assume temporarily  $\delta = 0$ . Under generic condition (**H2.1**, **H2.2**) on  $V$  and applying Theorem 3.4, there is a manifold  $\Pi_{n-1}^0$  which is normally hyperbolic and invariant for  $\Phi_{H''}^t$  and diffeomorphic to  $\mathbb{T}^{n-1} \times \mathbb{R}^{n-1}$ . As this system is a direct product of mechanical system and an integrable one, for each  $c \in \text{int } \mathcal{L}_{\beta_{H''}}(\omega_1, 0, \hat{\omega}_{n-2})$  the Mañé set stays in  $\mathcal{L}_{H''}(\Pi_{n-1}^0)$ . The manifold undergoes slight deformation  $\Pi_{n-1}^0 \rightarrow \Pi_{n-1}^\delta$  for small  $\delta > 0$ . For the same reason, the Mañé set for  $c \in \text{int } \mathcal{L}_{\beta_{H''}}(\omega_1, 0, \hat{\omega}_{n-2})$  also entirely stays in  $\Pi_{n-1}^\delta$  for small  $\delta > 0$ .

**Lemma 5.1.** *For  $C^2$ -Hamiltonian  $H(\tilde{x}, \tilde{y})$  defined on  $T^*\mathbb{T}^2$  and the first cohomology class  $c^* \in H^1(\mathbb{T}^2, \mathbb{R})$ , we assume that the Mather set  $\tilde{\mathcal{M}}(c^*)$  is supported on a  $\lambda g$ -minimal hyperbolic periodic orbit with  $g \in H_1(\mathbb{T}^2, \mathbb{Z})$ . Then, the set  $\mathcal{L}_{\beta_L}(\lambda g)$  (Fenchel-Legendre transformation of  $\lambda g$ ) is an interval  $\{c^* + [c_-, c_+]c_g\} \subset H^1(\mathbb{T}^2, \mathbb{R})$  with*



$c_- < c_+$  and  $\|c_g\| = 1$  such that for each  $c \in \{c^* + (c_-, c_+)c_g\}$  we have

$$\tilde{\mathcal{A}}(c) = \tilde{\mathcal{M}}(c^*).$$

*Proof.* The proof of this lemma is a variant of Theorem 3.1 of [C12]. As the system is autonomous with two degrees of freedom,  $\mathcal{L}_{\beta_H}(\lambda g)$  is either an interval or a point. In the case of interval, some  $c_g \in H^1(\mathbb{T}^2, \mathbb{R})$  exists such that  $\mathcal{L}_{\beta_H}(\lambda g) = \{c^* + [c_-, c_+]c_g\}$ . It follows from [Ms] that for all classes  $c \in \{c^* + (c_-, c_+)c_g\}$ , the Aubry sets  $\tilde{\mathcal{A}}(c)$  are the same. Let us show that  $c_- < c_+$  and  $\tilde{\mathcal{A}}(c) = \tilde{\mathcal{M}}(c^*)$  for  $c \in \{c^* + (c_-, c_+)c_g\}$ .

Given any absolutely continuous curve  $\gamma$ , its Lagrange action is defined as follows

$$A_c(\gamma) = \int L_H(\dot{\gamma}, \gamma) - \eta_c + \alpha_H(c) dt, \quad [\eta_c] = c.$$

Denote by  $\gamma_0$  the hyperbolic periodic orbit, we consider *minimal* homoclinic orbits to  $\gamma_0$ , which is located in the intersection of the stable and unstable manifolds of  $(\dot{\gamma}_0, \gamma_0)$ . A homoclinic orbit  $(\dot{\gamma}, \gamma)$  is called *minimal* if the lift of  $\gamma$ ,  $\tilde{\gamma}: \mathbb{R} \rightarrow \tilde{M}$  is semi-static for the class  $c^*$ , where  $\tilde{M}$  is the largest covering space of  $\mathbb{T}^2$  so that  $\pi_1(\tilde{M}) = \pi_1(U)$  holds for each open neighborhood of  $\mathcal{M}(c^*)$ . Because of the topology of  $\mathbb{T}^2$ , there are only two types of minimal homoclinic orbits, denoted by  $(\dot{\gamma}^\pm, \gamma^\pm)$ . Given a point  $x \in \gamma_0$ , there are four sequences of time  $t_{i,\pm}^\pm$  such that  $\gamma^-(t_{i,-}^\pm) \rightarrow x$  as  $t_{i,-}^\pm \rightarrow \pm\infty$  and  $\gamma^+(t_{i,+}^\pm) \rightarrow x$  as  $t_{i,+}^\pm \rightarrow \pm\infty$ ,  $t_{i,+}^\pm \rightarrow \pm\infty$  and  $t_{i,-}^\pm \rightarrow \pm\infty$  as  $i \rightarrow \infty$ . We define

$$A_c(\gamma^-, x) = \liminf_{i \rightarrow \infty} \int_{t_{i,-}^-}^{t_{i,-}^+} \left( L_H(\dot{\gamma}^-, \gamma^-) - \langle c, \dot{\gamma}^- \rangle + \alpha_H(c) \right) dt$$

$$A_c(\gamma^+, x) = \liminf_{i \rightarrow \infty} \int_{t_{i,+}^-}^{t_{i,+}^+} \left( L_H(\dot{\gamma}^+, \gamma^+) - \langle c, \dot{\gamma}^+ \rangle + \alpha_H(c) \right) dt$$

We obviously have  $A_{c^*}(\gamma^\pm, x) \geq 0$ . Next, we claim that

$$A_{c^*}(\gamma^+, x) + A_{c^*}(\gamma^-, x) > 0.$$

Otherwise, we would have  $A_{c^*}(\gamma^\pm) = 0$  for both  $\pm$ , which implies that  $\gamma^\pm \subset \tilde{\mathcal{A}}(c^*)$ . However, this violates the graph property of the Aubry set since  $[\gamma^+] \neq [\gamma^-]$  in the first relative homology group  $H_1(\mathbb{T}^2, \gamma_0, \mathbb{Z})$ , the projections of  $\gamma^\pm$  on  $\mathbb{T}^2$  must intersect. The contradiction proves our claim. Let us assume  $A_{c^*}(\gamma^+) > 0$  without loss of generality.

Since the minimal measure for class  $c^*$  is supported only on the periodic orbit,  $\tilde{\mathcal{N}}(c^*)$  is composed of those minimal homoclinic orbits along which the action equals zero. According to the upper semi-continuity of Mañé set in cohomology class, any minimal measure  $\mu_c$  is supported by a set lying in a small neighborhood of these homoclinic orbits if  $c = c^* + \Delta c$  and  $|\Delta c|$  is very small. Therefore, it is possible that  $\mu_c$  is supported in a neighborhood of  $\gamma^-$  only when  $A_{c^*}(\gamma^-) = 0$ .

For a  $\Delta c \neq 0$  so that  $\langle \Delta c, [\gamma_0] \rangle = 0$ , we get  $\langle \Delta c, [\gamma^+] \rangle > 0$  and  $\langle \Delta c, [\gamma^-] \rangle < 0$  by changing  $\Delta c$  to  $-\Delta c$ . We claim that the minimal measure  $\mu_c$  for  $c = c^* + \Delta c$  is still supported on this periodic orbit. Indeed, if  $\mu_c$  is not supported on that periodic orbit, its support lies in the small neighborhood of  $\gamma^-$ , it follows that  $-\langle \Delta c, \rho(\mu_c) \rangle > 0$ . On the other hand, as the  $c^*$ -minimal measure is uniquely supported on the periodic orbits, the  $\alpha$ -function is differentiable at  $c^*$  and  $\rho(\mu_{c^*}) = \lambda[\gamma_0]$  hold for certain number  $\lambda$ . Therefore, we have  $\alpha_H(c^* + \Delta c) - \alpha_H(c^*) = O(|\Delta c|^2)$ . Consequently, we obtain

from the definition that

$$\begin{aligned} A_c(\mu_c) &= \int (L_H - \eta_{c^*}) d\mu_c + \alpha_H(c^* + \Delta c) - \langle \Delta c, \rho(\mu_c) \rangle \\ &= \int (L_H - \eta_{c^*}) d\mu_c + \alpha_H(c^*) - \langle \Delta c, \rho(\mu_c) \rangle + O(|\Delta c|^2), \end{aligned}$$

from which we have  $A_c(\mu_c) > 0$  as  $A_{c^*}(\mu_c^*) \geq 0$ ,  $-\langle \Delta c, \rho(\mu_c) \rangle > 0$  and  $O(|\Delta c|^2)$  is a higher order term of  $|\Delta c|$ . The contradiction implies that  $\mathcal{L}_{\beta_H}(\lambda g)$  is an interval, not a point.  $\square$

Along the admissible resonance line  $\omega_a^\sharp = \lambda_a^\sharp(a, \frac{p_2}{q_2}, \frac{p_3}{q_3}, \dots, \frac{p_n}{q_n})$  in (4.8), we encounter finitely many strong double resonance points. Once a strong double resonance appear during the  $j^{th}$  step of reduction, it continues to be strong double resonance point for all the later steps. We remove an  $\varepsilon^\sigma$  interval centered around each strong double resonance point. We denote by  $a_-''$  and  $a_+''$  two consecutive strong double resonance points. Then we get an interval  $[a_-, a_+]$  satisfying  $|a_- - a_-''| = |a_+ - a_+''| = \varepsilon^\sigma$ ,  $\sigma < 1/2$ .

With the normal form of the Hamiltonian systems in (4.9) and (4.21) after all the reduction of orders, we have the following description of the Aubry sets.

**Theorem 5.1** (Description of the Aubry set). *Assume (H1)<sub>j</sub>, (H2)<sub>j</sub> for the Hamiltonian (1.1) during each step of reduction of order. Along the admissible resonance line  $\omega_a^\sharp$  there are finitely many strong double resonance points we have the following description of Aubry sets for the nonautonomous system along the admissible complete resonant line  $\omega_a^\sharp$ .*

- (1) *With finitely many exceptional frequencies  $\omega_{a^*}^\sharp$ , where we denote  $a^*$  for a typical one, the set  $\mathcal{L}_\beta(\omega_a^\sharp)$ ,  $a \in [a_-, a_+]$ , is a  $(n-1)$ -dimensional connected set that is homeomorphic to  $[0, 1]^{n-1}$ . For each  $c \in \text{int} \mathcal{L}_\beta(\omega_a^\sharp)$ , the Aubry set  $\mathcal{A}(c)$  lies in a four dimensional normally hyperbolic invariant cylinders.*
- (2) *For the finitely many exceptions, and for each  $c \in \text{int} \mathcal{L}_\beta(\omega_{a^*}^\sharp)$ , the Aubry set  $\mathcal{A}(c)$  lies in two distinct four dimensional normally hyperbolic invariant cylinders.*
- (3) *The above two items hold also for  $a$  satisfying  $|a - a''| \in [\gamma_{a''} \varepsilon^{1/2}, \varepsilon^\sigma]$  where  $\gamma_{a''} := \frac{\partial h}{\partial I}(\gamma)$  for  $\gamma$  defined in Remark 3.2 and to be determined in Section 6.3.*

*Proof.* After  $n-3$  steps of reduction of order, in Proposition 4.1, we arrive at Hamiltonian systems of three degrees of freedom possessing a four dimensional normally hyperbolic invariant cylinder. We know that the Aubry set of the original system lies inside the cylinder. Restricting to the cylinder, picking a Poincaré section, we get that the time-1 map is a twist map on a cylinder of dimension 2 and the  $\omega_a^\sharp$  is the rotation vector of the Aubry-Mather sets of the twist map. The description of the set  $\mathcal{L}_\beta(\omega_a^\sharp)$  in item (1) follows from repeated application of Lemma 5.1.  $\square$

**5.2. The flat  $\mathbb{F}_0$  of the  $\alpha$ -function.** From now on we study a neighborhood of the complete resonance.

**5.2.1. The Hamiltonian systems.** Recall the systems (4.19) and (4.20)

$$(5.1) \quad Y_\delta = \frac{1}{2} \langle Ay, y \rangle + \sum_{j=2}^{n-1} \delta_j V_j(x_1, \dots, x_i) + \delta_n R(x, x_n, y)$$

where  $x = (x_1, \dots, x_{n-1})$ ,  $y = (Y_1, \dots, Y_{n-1}) \in T^*\mathbb{T}^{n-1}$  and

$$(5.2) \quad Y_\delta := \mathfrak{S}_{II,2}^{-1*} Y_\delta = \langle \tilde{A}\tilde{y}, \tilde{y} \rangle + V_2(\tilde{x}) + \sum_{i=3}^{n-1} \left( \frac{b_i}{2} y_i^2 + \delta_i V_i(\tilde{x}_i) \right) + \delta_n R,$$

defined on  $T^*\mathbb{T}^2 \times \mathbb{R}^{2(n-3)} \times \mathbb{T}^1$ , where

$$(5.3) \quad V_i(\tilde{x}_i) = V_i(S_{II}^{-1}(\tilde{x}_i, \hat{0}_{n-i-1})), \quad \tilde{x} = (x_1, x_2), \quad \tilde{x}_i = (x_1, \dots, x_i)$$

and similarly for  $\tilde{y}, \tilde{y}_i$ . The matrix  $S_{II} \in SL(n-1, \mathbb{R})$  is given in (4.14) with  $j = n-1$  and the last row and column removed. We let  $\delta_2 = 1$ . The system  $Y_\delta$  is defined on  $T^*\mathbb{T}^{n-1} \times \mathbb{T}^1$  and  $Y_\delta$  is defined on  $T^*\mathbb{T}^2 \times \mathbb{R}^{2(n-3)} \times \mathbb{T}^1$ . The precise form of the term  $\delta_n R$  is not involved into the construction of diffusing orbit around the complete resonance provided it is small enough, so we do not bother to keep track of it under the transformations.

We introduce the following subsystem of  $Y_\delta$

$$(5.4) \quad \tilde{G}(\tilde{x}, \tilde{y}) = \frac{1}{2} \langle \tilde{A}\tilde{y}, \tilde{y} \rangle + V_2(\tilde{x}), \quad (\tilde{x}, \tilde{y}) \in T^*\mathbb{T}^2.$$

In  $Y_\delta$ , when  $\delta_i \rightarrow 0$ , the coefficient  $b_i = \text{const.} |\mathbf{k}^i|^2 \rightarrow \infty$  where  $\mathbf{k}^i$  is the  $i$ -th resonant integer vector depending on  $\delta_i$  (see (3.49) and (4.15) for the estimate of  $b_i$  and Section 4.2.1 and 4.2.2 for the definition of  $\mathbf{k}_i$ ). To resolve this singular behavior, we make the following rescaling to get a new Hamiltonian defined on  $T^*\mathbb{T}^2 \times \mathbb{R}^{2(n-3)} \times \mathbb{T}^1$  denoted by  $\mathbf{Y}_\delta$

$$(5.5) \quad \sqrt{b_i} y_i := y_i, \quad \frac{1}{\sqrt{b_i}} x_i := x_i, \quad i \geq 3, \quad \tilde{y} = y, \quad \tilde{x} = x,$$

$$(5.6) \quad \mathbf{Y}_\delta = \langle \tilde{A}\tilde{y}, \tilde{y} \rangle + V_2(\tilde{x}) + \sum_{i=3}^{n-1} \left( \frac{1}{2} y_i^2 + \delta_i \mathbf{V}_i(\tilde{x}_i) \right) + \delta_n R,$$

where  $\mathbf{V}_i(\tilde{x}, x_3, \dots, x_i) = V_i(\tilde{x}, \sqrt{b_3} x_3, \dots, \sqrt{b_i} x_i)$ . We introduce a matrix

$$\mathbf{S} = \text{diag} \left\{ 1, 1, \frac{1}{\sqrt{b_3}}, \dots, \frac{1}{\sqrt{b_{n-1}}} \right\} S_{II} \in GL(n-1, \mathbb{R})$$

So that  $Y_\delta$  and  $\mathbf{Y}_\delta$  are related via the change of coordinates  $\mathbf{x} = \mathbf{S}x$ ,  $\mathbf{y} = \mathbf{S}^{-t}y$ . The auxiliary system  $\mathbf{Y}_\delta$  is easier to study than  $Y_\delta$ . In the following, we always study system  $\mathbf{Y}_\delta$  first, then transform the results to that of  $Y_\delta$  using  $\mathbf{S}$ .

However, these transformations may lead to significant change of the norm of the potentials, which we estimate in the following. Notice that  $\delta_i V_i$  consists of Fourier modes of  $V$  in (2.13) in the span  $\{\mathbf{k}^1, \dots, \mathbf{k}^i\}$  with  $|\mathbf{k}^1|, |\mathbf{k}^2| \ll |\mathbf{k}^3| \ll \dots \ll |\mathbf{k}^{n-1}|$  (see (4.11) in Section 4.2.1). Since we assume the Hamiltonian system is in  $C^r$ ,  $r > 2n$ , we get that the estimate  $\|\delta_i V_i\|_{C^2} \leq \frac{C}{|\mathbf{k}^i|^{2n-2}}$ . From the definition of  $\mathbf{V}_i$  in (5.3), we see that only the first  $i$  rows of  $S_{II}$  enters  $\mathbf{V}_i$  through composing with  $V_i$ . Since the estimate of the  $i$ -th row of  $S_{II}$  is the same as that of  $s_i = \text{const.} |\mathbf{k}^i|$ , we get that  $\|\delta_i \mathbf{V}_i\|_{C^2} \leq \frac{C}{|\mathbf{k}^i|^{2n-4}}$ . Since  $S_{II}$  in (4.14) is lower triangular with diagonal entries 1, we get that the dependence of  $\mathbf{V}_i$  on  $\tilde{x}_i$  is the same as that of  $V_i$  on  $\tilde{x}_i$ . This implies that  $\|\delta_i \mathbf{V}_i\|_{C^2} \leq \frac{C}{|\mathbf{k}^i|^{2n-4}}$ .

So we get that if we let  $\delta_i \rightarrow 0$ ,  $i > 3$ , in  $\mathbf{Y}_\delta$ , the potentials  $\delta_i \mathbf{V}_i$  go to zero in the  $C^2$  norm.

5.2.2. *The  $\alpha$ -function and its flat.* The  $\alpha$ -function  $\alpha_{Y_\delta}$  of  $Y_\delta$  is defined in the appendix. We denote the flat of  $\alpha_{Y_\delta}$  by

$$\mathbb{F}_0 := \arg \min \alpha_{Y_\delta}.$$

By Theorem 3.1 of [C12] (Theorem 3.3 here) we know that  $\mathbb{F}_0$  is an  $(n-1)$  dimensional convex set. However, this is far from being enough for us to handle the problem of crossing  $m$ -strong resonance. We need more information about the shape of the flat  $\mathbb{F}_0$ . It is not straightforward to study the flat  $\mathbb{F}_0$ . Instead, we study it through studying the flat for  $Y_\delta$  and  $\mathbf{Y}_\delta$ .

As for the  $\alpha$ -functions of  $Y_\delta$  and  $\mathbf{Y}_\delta$ , since the Lagrangians are not defined on  $T\mathbb{T}^{n-1} \times \mathbb{T}$ , it is not given by the standard definition in Mather theory. We simply define the  $\alpha$ -function of  $Y_\delta$ ,  $\mathbf{Y}_\delta$  as the linear transformation of  $\alpha_{Y_\delta}$  as follows

$$\alpha_{Y_\delta}(\mathbf{c}) := \alpha_{Y_\delta}(S_{II}^t \mathbf{c}), \quad \alpha_{\mathbf{Y}_\delta}(\mathbf{c}) := \alpha_{Y_\delta}(\mathbf{S}^t \mathbf{c}).$$

We denote the corresponding flat for  $Y_\delta$  and  $\mathbf{Y}_\delta$  by  $\mathbb{F}_0$  and  $\mathbf{F}_0$  respectively.

To study the flats of the  $\alpha$  functions, we introduce three systems obtained from setting  $\delta_i = 0, \forall i \geq 3$  in  $Y_\delta, \mathbf{Y}_\delta$  respectively. Notice that  $Y_0, \mathbf{Y}_0$  are artificial since when  $\delta_i \rightarrow 0$ , we have  $b_i \rightarrow \infty$ . However, we artificially set  $\delta_i \rightarrow 0, i \geq 3$  but leave  $b_i$  fixed. We only use these artificially created systems to assist us to understand the structure of  $\alpha$  functions, but not use them to construct diffusing orbits. The system  $\mathbf{Y}_0$  is well defined in  $C^2$  topology by taking limit  $\delta_i \rightarrow 0, i \geq 3$ , in  $\mathbf{Y}_\delta$  directly.

We denote by  $\tilde{\alpha} : H^1(\mathbb{T}^2, \mathbb{R}) \rightarrow \mathbb{R}$  the  $\alpha$ -function for the  $\tilde{\mathbf{G}}$ .

**Lemma 5.2.** *For the systems  $Y_0, \mathbf{Y}_0$ , we have the following expressions for their  $\alpha$  functions.*

$$\alpha_{Y_0}(\mathbf{c}) = \tilde{\alpha}(\tilde{\mathbf{c}}) + \sum_{i=3}^{n-1} \frac{b_i}{2} \mathbf{c}_i^2, \quad \alpha_{\mathbf{Y}_0}(\mathbf{c}) = \tilde{\alpha}(\tilde{\mathbf{c}}) + \frac{1}{2} \|\tilde{\mathbf{c}}\|_{\ell_2}^2, \quad \alpha_{Y_0} = S_{II}^{-t*} \alpha_{\mathbf{Y}_0} = \mathbf{S}^{-t*} \alpha_{\mathbf{Y}_0}.$$

*Proof.* We only need to prove the expression for  $\alpha_{Y_0}$  and obtain the other two by linear transformations. From its definition

$$\begin{aligned} -\alpha_{Y_0}(c) &= \inf_{\mu} \int L(\dot{x}, x) d\mu = \inf_{\mu} \int \langle A^{-1} \dot{x}, \dot{x} \rangle + V_2(\tilde{x}) - \langle c, \dot{x} \rangle d\mu \\ &= \inf_{\mu} \int \langle S_{II}^{-t} A^{-1} S_{II}^{-1} \dot{x}, \dot{x} \rangle + V_2(\tilde{x}) - \langle \mathbf{c}, \dot{x} \rangle d\mu \\ &= \inf_{\mu} \int \langle \tilde{A}^{-1} \dot{\tilde{x}}, \dot{\tilde{x}} \rangle + V_2(\tilde{x}) - \langle \tilde{\mathbf{c}}, \dot{\tilde{x}} \rangle + \sum_{i=3}^{n-1} \frac{b_i^{-1}}{2} \dot{\tilde{x}}_i^2 - \langle \tilde{\mathbf{c}}, \dot{\tilde{x}} \rangle d\mu \\ (5.7) \quad &= \inf_{\mu} \int \langle \tilde{A}^{-1} \dot{\tilde{x}}, \dot{\tilde{x}} \rangle + V_2(\tilde{x}) - \langle \tilde{\mathbf{c}}, \dot{\tilde{x}} \rangle d\mu + \inf_{\mu} \int \sum_{i=3}^{n-1} \frac{b_i^{-1}}{2} \dot{\tilde{x}}_i^2 - \langle \tilde{\mathbf{c}}, \dot{\tilde{x}} \rangle d\mu \\ &= \tilde{\alpha}(\tilde{\mathbf{c}}) + \sum_{i=3}^{n-1} \frac{b_i}{2} \mathbf{c}_i^2 \end{aligned}$$

where in the second equality, we use  $S_{II} \dot{x} = \dot{\tilde{x}}, S_{II}^{-t} c = \mathbf{c} = (\tilde{\mathbf{c}}, \hat{\mathbf{c}})$  while leaving  $\tilde{x}$  unchanged. Going back to the  $c$  variables, we get the expression of  $\alpha_{Y_0}$ .  $\square$

First we get that the flat  $\mathbb{F}_0$  and  $\mathbf{F}_0$  for the system  $Y_0, \mathbf{Y}_0$  are a two-dimensional disk  $\tilde{\mathbb{F}}_0 \times \{\hat{\mathbf{c}} = \hat{0}\}$ , where  $\tilde{\mathbb{F}}_0$  is the flat of  $\tilde{\alpha}$ .

The next lemma describes the flat  $\mathbf{F}_0$  of the system  $\mathbf{Y}_\delta$ . It is not hard to get a description of the flat  $\mathbf{F}_0$  and  $\mathbb{F}_0$  after linear transformations.

**Proposition 5.1.** *For  $C^r$ ,  $r \geq 2$  generic  $V_i$ ,  $i = 2, \dots, n-1$ , the flat  $\mathbf{F}_0$  of the system  $\mathbf{Y}_\delta$  is  $(n-1)$  dimensional and entirely stays in  $O(\sqrt{\delta_3})$ -neighborhood of the disk  $\tilde{\mathbb{F}}_0 \times \{\hat{\mathbf{c}} = \hat{0}\}$ . More precisely, the width of the cube in the  $\mathbf{c}_i$  direction is  $O(\sqrt{\delta_i})$ ,  $i = 3, 4, \dots, n-1$ .*

*Proof.* The fact that  $\mathbf{F}_0$  is  $n-1$  dimensional is given by Theorem 3.1 of [C12] (Theorem 3.3 here). Since we have  $|Y_\delta - Y_0| < \delta_3$ , we have (cf. [C11])

$$|\alpha_{Y_\delta}(c) - \alpha_{Y_0}(c)| \leq \delta_3, \quad \forall c \in H^1(\mathbb{T}^{n-1}, \mathbb{R}).$$

After linear transformations, this gives

$$(5.8) \quad |\alpha_{\mathbf{Y}_\delta}(\mathbf{c}) - \alpha_{\mathbf{Y}_0}(\mathbf{c})| \leq \delta_3.$$

Since we have  $\alpha_{\mathbf{Y}_0}(\mathbf{c}) = \tilde{\alpha}(\tilde{\mathbf{c}}) + \frac{1}{2}\|\tilde{\mathbf{c}}\|_{\ell_2}^2$ , we get that

$$\alpha_{\mathbf{Y}_0}(\tilde{\mathbf{c}}, \hat{\mathbf{c}}) > \frac{1}{2}K^2\delta_3, \text{ if } |\hat{\mathbf{c}}| > K\sqrt{\delta_3}.$$

As  $\tilde{\alpha}$  is non-negative, it follows from Formula (5.8) that  $\alpha_{\mathbf{Y}_\delta}(\tilde{\mathbf{c}}, \hat{\mathbf{c}}) > (\frac{1}{2}K^2 - 1)\delta_3$ . Also due to Formula (5.8), we have  $\min \alpha_{\mathbf{Y}_\delta} \leq \delta_3$ . Therefore,

$$\alpha_{\mathbf{Y}_\delta}(\tilde{\mathbf{c}}, \hat{\mathbf{c}}) > \min \alpha_{\mathbf{Y}_\delta}, \text{ if } |\hat{\mathbf{c}}| > 2\sqrt{\delta_3}.$$

This completes the proof for the  $O(\sqrt{\delta_3})$  estimate. To get the  $O(\sqrt{\delta_i})$  estimate for  $i > 3$ , we only need to set  $\delta_j = 0$ ,  $j > i$  while keeping  $\delta_j > 0$ ,  $j < i$  and repeat the above argument.  $\square$

Therefore, the flat  $\mathbf{F}_0$  looks like a pizza, horizontal in the direction of  $\tilde{\mathbf{c}}$  with small thickness of order  $O(\sqrt{\delta_3})$ .

**5.3. The channel connected to  $\mathbb{F}_0$ .** Our diffusing orbit is constructed along a NHIC of dimension 4 (for the Hamiltonian flow, which can be reduced to dimension 2 for the Poincaré map) until we get close to a complete resonance. This NHIC corresponds to a channel in  $H^1(\mathbb{R}^{n-1}, \mathbb{R})$  according to item (1) of Theorem 5.1. Let us describe this channel.

In the KAM normal form (4.11) with  $j = n-3$  and  $j+2 = n-1$ , we have the frequency segment  $\omega_a^\sharp = \lambda_a^\sharp(a, \frac{p_2}{q_2}, \dots, \frac{p_n}{q_n})$  as the rotation vector of the Aubry-Mather set on a NHIC of dimension 4 (for the Hamiltonian flow, which can be reduced to dimension 2 for the Poincaré map). At complete resonance  $a$  is also rational. Applying the symplectic transformation  $\mathfrak{M}_{n-1}$  in Section 4.2.2 and performing the energetic reduction, we obtain for the system  $Y_\delta$  the frequency  $\lambda \mathbf{e}_1$  along the cylinder and  $\lambda = 0$  corresponds to the complete resonance.

We define the channel  $\mathbb{C}$  associated to the system  $Y_\delta$  as follows

$$(5.9) \quad \mathbb{C} = \bigcup_{\lambda \neq 0} \mathcal{L}_{\beta_{Y_\delta}}(\lambda \mathbf{e}_1).$$

We want to see how this channel is connected to the flat  $\mathbb{F}_0$ . The channel  $\mathbb{C}$  transforms to that associated to the system  $\mathbf{Y}_\delta$ ,  $\mathbf{Y}_\delta$  via linear transformations  $S_{II}^t$ ,  $\mathbf{S}^t$  respectively. We have the following description of the channel for the system  $\mathbf{Y}_\delta$ . Again it is not hard to get the corresponding description for the system  $Y_\delta$  after the linear transformation  $\mathbf{S}^t$ .

**Lemma 5.3.** *The channel associated to the system  $\mathbf{Y}_\delta$  is a neighborhood of*

$$\left(\tilde{\mathbb{C}}, \frac{s_3}{\sqrt{b_3}}\lambda, \dots, \frac{s_{n-1}}{\sqrt{b_{n-1}}}\lambda\right),$$

where  $\tilde{\mathbb{C}} = \bigcup_{\lambda \neq 0} \mathcal{L}_{\beta_{\tilde{\mathbb{C}}}}(\lambda, 0)$  is a two dimensional channel for the system  $\tilde{\mathbf{G}}$  corresponding to the NHIC obtained by Theorem 3.4 with homology class  $g = (1, 0)$ . The size of the neighborhood projected to the  $\mathbf{c}_i$  component has width  $\sqrt{\delta_i}$ . Moreover, the flat as well as the two arms of the channel with  $\lambda > 0$  and  $\lambda < 0$  respectively are centrally symmetric as  $\delta_n \rightarrow 0$ .

*Proof.* Again we look at system  $\mathbf{Y}_\delta$  and  $\mathbf{Y}_\delta$ . In system  $\mathbf{Y}_\delta$ , the frequency becomes

$$S_{II}\lambda\mathbf{e}_1 = \lambda(1, 0, s_3, \dots, s_{n-1}).$$

In system  $\mathbf{Y}_\delta$ , the frequency is

$$\mathbf{S}\lambda\mathbf{e}_1 = \lambda\left(1, 0, \frac{s_3}{\sqrt{b_3}}, \dots, \frac{s_{n-1}}{\sqrt{b_{n-1}}}\right).$$

Next, we set  $\delta_i = 0$ ,  $i \geq 3$  in  $\mathbf{Y}_\delta$  to work with the system  $\mathbf{Y}_0$ . Performing the Legendre transform we get the following two dimensional channel for  $\mathbf{Y}_0$

$$\left(\tilde{\mathbb{C}}, \frac{s_3}{\sqrt{b_3}}\lambda, \dots, \frac{s_{n-1}}{\sqrt{b_{n-1}}}\lambda\right).$$

When we turn on the potentials in  $\mathbf{Y}_\delta$  and  $\mathbf{Y}_\delta$ , this channel becomes  $(n-1)$  dimensional according to Theorem 5.1. Applying the same argument of Proposition 5.1 to the function  $\alpha_{\mathbf{Y}_\delta}(\mathbf{c}) - \langle \lambda\mathbf{e}_1, \mathbf{c} \rangle$  instead of  $\alpha_{\mathbf{Y}_\delta}$  we get that for all  $\mathbf{c} \in \mathbb{C}$ , there exists  $\lambda$  such that

$$\left|\mathbf{c}_i - \frac{s_3}{\sqrt{b_3}}\lambda\right| \leq \sqrt{\delta_i}, \quad i \geq 3.$$

So we get the description of the channel associated to the system  $\mathbf{Y}_\delta$ . When transformed to the system  $\mathbf{Y}_\delta$ , the channel is a neighborhood of  $(\tilde{\mathbb{C}}, \frac{s_3}{b_3}\lambda, \dots, \frac{s_{n-1}}{b_{n-1}}\lambda)$ , whose projection to the  $c_i$  component has width  $\sqrt{\delta_i/b_i}$ .

Moreover, when we set  $\delta_n = 0$  in all the three systems  $\mathbf{Y}_\delta, \mathbf{Y}_\delta, \mathbf{Y}_\delta$ , we get that the  $\alpha$  function is an even function hence the flat  $\mathbb{F}_0, \mathbf{F}_0, \mathbf{F}_0$  as well as the two arms of the channel for  $\lambda > 0$  and  $\lambda < 0$  are centrally symmetric.  $\square$

## 6. HOW TO CROSS $m$ -STRONG RESONANCES

**6.1. The difficulty and the strategy.** For nearly integrable systems, it is unavoidable to cross multiple-strong resonance if we want to find global diffusion orbits.

As it is unclear to us if the channel  $\mathbb{C}$  is connected to the flat  $\mathbb{F}_0$  or not, we avoid entering the flat along channel. Instead, the way we find stays away from the  $m$ -strong resonance by a distance  $\gamma\varepsilon^{1/2}$  (see Remark 3.2) and turns around the flat so that two channels are connected. It is for two reasons: the singular behavior (Lemma 3.4) of the action-angle coordinates, and the lack of the regularity of the Hamiltonian system restricted on the NHIC close to the hyperbolic fixed point applying Theorem 3.2 which gives us only a  $C^1$  Hamiltonian since  $\ln \lambda / |\ln \mu|$  in Theorem 3.2 can be close to one.

For three degrees of freedom system, the first named author discovered in [C12] a path of  $c$ -equivalence connecting to arms of the channel turning around but without

entering the flat  $\mathbb{F}_0$ . This  $c$ -equivalence mechanism is essentially a diffusing mechanism for systems of two degrees of freedom close to the hyperbolic fixed point.

When we apply the mechanism to systems with more than three degrees of freedom, we find that the path in  $H^1(\mathbb{T}^{n-1}, \mathbb{R})$  has constant last  $n-3$  components  $\hat{c}$  for  $Y_\delta, Y_\delta, \mathbf{Y}_\delta$  since the upper triangular matrices  $S_{II}^{-t}, \mathbf{S}^{-t}$  preserves the constancy of the last  $n-3$  entries. If two channels are connected to the flat with different “high”, they can not be connected by a path of  $c$ -equivalence.

For instance, if we move the frequency from  $\lambda e_1 \rightarrow -\lambda e_1$ , correspondingly, we send  $c = (\tilde{c}, \hat{c}) \rightarrow -c = (-\tilde{c}, -\hat{c})$ . Due to the central symmetry, it is allowed that the image lies in a neighborhood of  $-c$  provided the neighborhood lies entirely in the channel  $\mathbb{C}$ . We consider the system  $\mathbf{Y}_\delta$  for which we need to send  $\mathbf{c} \in \mathbb{C}$  to a neighborhood in  $\mathbb{C}$  of  $-\mathbf{c}$ . A new difficulty arises (see Figure 2). The two arms of the  $\mathbb{C}$  of  $s$ -resonance reach the flat  $\mathbf{F}_0$  at different “height” due to the symmetry, and they are not connected by path of cohomology equivalence used in [C12], along which the quantity  $\hat{c}$  remains unchanged (contour line). Our strategy consists of two steps. The first step is to perform the cohomology equivalence by holding  $\hat{c} = \hat{c}^*$  constant and sending the first two components  $\tilde{c} \rightarrow -\tilde{c}$  (the red horizontal curve in Figure 2). So we get the cohomology  $(-\tilde{c}, \hat{c}^*)$  after the cohomology equivalence procedure. The second step is to send  $\hat{c}^*$  to  $-\hat{c}^*$ . To achieve this, we construct a short transition chain connect one arm of the channel to a contour line which intersects the other arm of the channel (the vertical curve in Figure 2). It looks like a ladder. So, starting from one NHIC we pass an  $m$ -strong resonance by following a curve of  $c$ -equivalence and climbing up to another NHIC by the ladder.

**6.2. Cohomology equivalence around  $m$ -strong resonance.** We assume that the strong double resonance is encountered at the first step of reduction, i.e. the  $(n-2)$ -strong resonance, other  $m$ -strong resonances are treated similarly.

**Theorem 6.1** (Theorem 5.1 of [C12]). *Let  $L_{\tilde{\mathbf{G}}} = \frac{1}{2}\langle \tilde{A}^{-1}\dot{\tilde{\mathbf{x}}}, \dot{\tilde{\mathbf{x}}} \rangle + V_2(\tilde{\mathbf{x}})$  be the Lagrangian corresponding to the Hamiltonian  $\tilde{\mathbf{G}}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$  via Legendre transform. A residual set  $\mathfrak{V} \subset C^r(\mathbb{T}^2, \mathbb{R})$  exists such that for each  $V_2 \in \mathfrak{V}$  and for each  $\tilde{c} \in \partial\tilde{\mathbb{F}}_0$ , where  $\tilde{\mathbb{F}}_0$  is the flat of the  $\alpha$  function for  $\tilde{\mathbf{G}}$ , the Mañé set does not cover the whole configuration space  $\mathbb{T}^2$ .*

*Sketch of the proof.* It is actually one of the main results in [C12], one can find the more detailed proof in Section 3 of [C15b]. Applying Theorem 3.3, we get a flat  $\tilde{\mathbb{F}}_0$  of the  $\alpha$ -function which is a 2-dimensional convex disk in  $H^1(\mathbb{T}^2, \mathbb{R})$ . Let  $\partial\tilde{\mathbb{F}}_0$  be the boundary of  $\tilde{\mathbb{F}}_0$ . The key step is showing that, under the hypothesis **(H2.1)**, there are at most four classes on  $\partial\tilde{\mathbb{F}}_0$  with which the Mañé set covers the whole configuration space  $\mathbb{T}^2$ . The rest of proof is easier, with extra perturbation, these four Mañé sets do not cover  $\mathbb{T}^2$  any more.  $\square$

Guaranteed by the upper semi-continuity of the Mañé set, we obtain the description of the structure of the Mañé set extends to energy levels slightly higher than  $\min \tilde{\alpha}$ .

**Proposition 6.1.** *Given  $C^r$ ,  $r > 2$ , generic  $V_2$ , some positive numbers  $\Delta_0 > 0$  exist, depending on  $V_2$ , so that for each  $E \in (0, \Delta_0)$  and each  $\tilde{c} \in \tilde{\alpha}^{-1}(E)$  there exists a circle  $\Sigma_{\tilde{c}} \subset \mathbb{T}^2$  so that all  $\tilde{c}$ -semi static curves of  $L_{\tilde{\mathbf{G}}}$  pass through that circle transversally and*

$$\mathcal{N}_{\tilde{\mathbf{G}}}(\tilde{c}) \cap \Sigma_{\tilde{c}} \subset \bigcup I_{\tilde{c},i}$$

where  $I_{\tilde{c},i} \subset \Sigma_{\tilde{c}}$  are closed intervals, disjoint to each other.

*Proof.* By the upper-semi continuity of Mañé set in first cohomology and the compactness of  $\partial\tilde{\mathbb{F}}_0$ , ceratin number  $\Delta_0 > 0$  exists such that  $\mathcal{N}(\tilde{c}) \subsetneq \mathbb{T}^2$  if  $\tilde{c} \in \tilde{\alpha}^{-1}(E)$  with  $E < \Delta_0$ .

The rotation vector of each minimal measure  $\mu_{\tilde{c}}$  is non-zero for any  $\tilde{c} \in \tilde{\alpha}^{-1}(E)$  if  $E > 0$ . As the configuration space is 2-torus, all  $\tilde{c}$ -semi static curves have to turn around the torus in the same direction. Therefore,  $\exists$  a circle intersects these curves topologically transversal. Since the Mañé set does not cover the 2-torus, the restriction of the Mañé set on this circle does not make up the whole circle.  $\square$

A cohomology class  $\mathbf{c} = (\tilde{c}, \hat{c})$  for the system  $\mathbf{Y}_\delta$  is transformed to  $c = (\tilde{c}, \hat{c})$  that of  $Y_\delta$  via  $\mathbf{c} = S_{II}^{-t}c$ . Next, a cohomology class  $\mathbf{c} = (\tilde{c}, \hat{c})$  for  $\mathbf{Y}_\delta$  is related to  $\mathbf{c} = (\tilde{c}, \hat{c})$  for  $\mathbf{Y}_\delta$  via a rescaling in (5.5). Since  $S_{II}^{-t}$  is upper triangular, a constant  $\hat{c}^*$  is transformed to a constant  $\hat{c}^*$ . Namely, the constancy of the last  $(n-3)$  entries of the cohomology classes is preserved by the  $S_{II}^{-t}$  transform.

Recall that  $\alpha_{\mathbf{Y}_0}(\mathbf{c}) = \tilde{\alpha}(\tilde{c}) + \frac{1}{2}\|\hat{c}\|_{\ell_2}^2$  in (5.2). Since we have  $\mathbf{Y}_\delta - \mathbf{Y}_0 = O(\delta_3)$ , there exists some constant  $d > 0$  such that

$$(6.1) \quad -d\delta_3 \leq \alpha_{\mathbf{Y}_\delta} - \alpha_{\mathbf{Y}_0} \leq d\delta_3.$$

When we look at the energy level  $E$  with  $E - \frac{1}{2}\|\hat{c}\|_{\ell_2}^2 \in (d\delta_3, \Delta_0 - d\delta_3)$ , the subsystem  $\tilde{\mathbf{G}}$  in  $\mathbf{Y}_\delta$  has energy in  $(0, \Delta_0)$ .

**Corollary 6.1.** *For any energy  $E$  with  $E - \frac{1}{2}\|\hat{c}^*\|_{\ell_2}^2 \in (d\delta_3, \Delta_0 - d\delta_3)$ , any  $(\tilde{c}, \hat{c}^*)$  so that  $\alpha_{\mathbf{Y}_\delta}(\tilde{c}, \hat{c}^*) = E$  and for sufficiently small  $\delta_3 \gg \delta_4 \gg \dots \gg \delta_n > 0$ , all semi static curves of  $L_{\mathbf{Y}_\delta}$  with cohomology class  $c = (\tilde{c}, \hat{c}^*)$  (related to  $\mathbf{c}(\tilde{c}, \hat{c}^*)$  via linear transformations) pass transversally through the section  $\Sigma_{\tilde{c}} \times \{\hat{x} \in \mathbb{T}^{n-3}\}$*

$$(6.2) \quad \mathcal{N}_{Y_\delta}(\tilde{c}, \hat{c}^*) \cap \left( \Sigma_{\tilde{c}} \times \{\hat{x} \in \mathbb{T}^{n-3}\} \right) \subset \bigcup I_{\tilde{c}, i} \times \{\hat{x} \in \mathbb{T}^{n-3}\}.$$

*Proof.* Let us consider the Mañé set  $\tilde{\mathcal{N}}_{Y_\delta}(\tilde{c}, \hat{c}^*)$  of  $Y_\delta$ , lift the last  $n-3$  components of points in the set to the universal covering space  $\mathbb{R}^{2(n-3)}$ . After the linear transformation relating  $Y_\delta$  and  $\mathbf{Y}_\delta$ , the Mañé set  $\tilde{\mathcal{N}}_{Y_\delta}(\tilde{c}, \hat{c}^*)$  is sent to an invariant set of  $\mathbf{Y}_\delta$  denoted by  $\tilde{\mathcal{N}}_{\mathbf{Y}_\delta}(\tilde{c}, \hat{c}^*)$  which has upper-semi-continuity as  $\delta_i \rightarrow 0$ ,  $i \geq 3$  in the Lagrangian system  $L_{\mathbf{Y}_\delta}$ . We have shown in Section 5.2.1 that the potentials  $\delta_i \mathbf{V}_i$  in  $\mathbf{Y}_\delta$  converge to zero in the  $C^2$  topology as  $\delta_i \rightarrow 0$ ,  $i \geq 3$ .

As the Hamiltonian flow  $\Phi_{\mathbf{Y}_0}^t$  is the direct product of  $\Phi_{\tilde{\mathbf{G}}}^t$  and an integrable system, the Mañé set  $\tilde{\mathcal{N}}_{\mathbf{Y}_0}(\tilde{c}, \hat{c}^*)$  of  $\mathbf{Y}_0$  is given by  $\tilde{\mathcal{N}}_{\tilde{\mathbf{G}}}(\tilde{c}) \times \{\hat{c}^*\} \times \mathbb{R}^{n-3}$ . So, we get that for sufficiently small  $\delta_3 \gg \delta_4 \gg \dots \gg \delta_n > 0$ ,

$$\mathcal{N}_{\mathbf{Y}_\delta}(\tilde{c}, \hat{c}^*) \cap \left( \Sigma_{\tilde{c}} \times \mathbb{R}^{n-3} \right) \subset \bigcup I_{\tilde{c}, i} \times \mathbb{R}^{n-3}.$$

Noticing that the matrix  $S_{II}$  is lower triangular and is  $\text{id}_2$  in the first  $2 \times 2$  block, we get that the projection of the two sets  $\tilde{\mathcal{N}}_{Y_\delta}(\tilde{c}, \hat{c}^*)$  and  $\tilde{\mathcal{N}}_{\mathbf{Y}_\delta}(\tilde{c}, \hat{c}^*)$  to the first  $\mathbb{T}^2$  factor are the same. For the same reason, the section  $\Sigma_{\tilde{c}} \subset \mathbb{T}^2$  remains unchanged under the linear transformations. This completes the proof.  $\square$

The set  $\Gamma_{\mathbf{Y}_0}(\hat{c}^*, E) = \{c \in H^1(\mathbb{T}^n, \mathbb{R}) : \hat{c} = \hat{c}^*, \alpha_{\mathbf{Y}_0}(\mathbf{c}) = E\}$  is homeomorphic to a circle in the plane  $H^1(\mathbb{T}^2, \mathbb{R}) \times \{\hat{c} = \hat{c}^*\}$ , encircling the disk  $\tilde{\mathbb{F}}_0 \times \{\hat{c}_k = \hat{c}^*\}$ . For any



$\mathbf{c}, \mathbf{c}' \in \Gamma_{\mathbf{Y}_0}(\hat{\mathbf{c}}^*, E)$ , we have  $\alpha_{\mathbf{Y}_0}(\mathbf{c}) = \alpha_{\mathbf{Y}_0}(\mathbf{c}')$  and  $\hat{\mathbf{c}} - \hat{\mathbf{c}}' = 0$ . Extending this property to the Lagrangian determined by  $\mathbf{Y}_\delta$  defined in (4.19), for  $E \geq \frac{1}{2}\|\hat{\mathbf{c}}^*\|_{\ell_2}^2 + d\delta_3$ , the set

$$(6.3) \quad \Gamma_{\mathbf{Y}_\delta}(\hat{\mathbf{c}}^*, E) = \{\mathbf{c} \in H^1(\mathbb{T}^{n-1}, \mathbb{R}) : \hat{\mathbf{c}} = \hat{\mathbf{c}}^*, \alpha_{\mathbf{Y}_\delta}(\mathbf{c}) = E\}$$

is a circle located in the 2-dimensional plane  $H^1(\mathbb{T}^2, \mathbb{R}) \times \{\hat{\mathbf{c}} = \hat{\mathbf{c}}^*\}$ , encircling the disk  $\tilde{\mathbb{F}}_0 \times \{\hat{\mathbf{c}} = \hat{\mathbf{c}}^*\}$ . For any two classes  $\mathbf{c}, \mathbf{c}'$  on the circle, one has

$$(6.4) \quad \hat{\mathbf{c}} - \hat{\mathbf{c}}' = 0, \quad \alpha_{\mathbf{Y}_\delta}(\mathbf{c}) = \alpha_{\mathbf{Y}_\delta}(\mathbf{c}').$$

We define

$$(6.5) \quad \mathbb{A}(\Delta_0, \hat{E}) = \bigcup_{\substack{\hat{\alpha}(\hat{\mathbf{c}}^*) \in [0, \hat{E}] \\ E \in [d\delta_3, \Delta_0]}} \Gamma_{\mathbf{Y}_\delta}(\hat{\mathbf{c}}^*, E)$$

which is an  $(n-1)$ -dimensional annulus. Each circle  $\Gamma_{\mathbf{Y}_\delta}(\hat{\mathbf{c}}^*, E)$  is transformed to a circle  $\Gamma_{Y_\delta}(\hat{c}^*, E)$  for the system  $Y_\delta$  after a linear transformation given by  $S_{II}$  and a rescaling, hence determines a circle in  $H^1(\mathbb{T}^n, \mathbb{R})$  (See Formula (4.18) for the definition of  $H_{n-1}$ ). We next define

$$\Gamma_{H_{n-1}}(\hat{c}^*, E) = \{(\tilde{c}, \hat{c}, c_n) : (\tilde{c}, \hat{c}) \in \Gamma_{Y_\delta}(\hat{c}^*, E), \quad c_n = \alpha_{Y_\delta}(\tilde{c}, \hat{c}^*)\}.$$

Due to the following theorem,  $\alpha_{H_{n-1}}(c, c_n) = E^*$  holds for all  $(c, c_n) \in \Gamma_{H_{n-1}}(\hat{c}^*, E)$ .

**Theorem 6.2** (Theorem 3.4 of [C12]). *For a Hamiltonian  $H(x, y, x_n, y_n)$  we assume that  $\partial_{y_n} H \neq 0$  on  $\{H^{-1}(E)\} \cap \{y_n \in [y_n^-, y_n^+]\}$ . Let  $y_n = Y_\delta(x, y, \tau)$  be the solution of  $H = E$  ( $\tau = -x_n$ ). Let  $\alpha_H$  and  $\alpha_{Y_\delta}$  be the  $\alpha$ -function for  $H$  and  $Y_\delta$  respectively, then for  $\alpha_{Y_\delta}(c) \in [y_n^-, y_n^+]$  we have  $(c, \alpha_{Y_\delta}(c)) \in \alpha_H^{-1}(E)$ .*

The  $c$ -equivalence to be set up is in the new version introduced in [LC]. The concept of the equivalence was first introduced in [M93], but it does not apply to interesting problems of autonomous system. We call  $\Sigma_c$  non-degenerately embedded  $(n-1)$ -dimensional torus by assuming a smooth injection  $\varphi: \mathbb{T}^{n-1} \rightarrow \mathbb{T}^n$  such that  $\Sigma_c$  is the image of  $\varphi$ , and the induced map  $\varphi_*: H_1(\mathbb{T}^{n-1}, \mathbb{Z}) \hookrightarrow H_1(\mathbb{T}^n, \mathbb{Z})$  is an injection.

Let  $\mathfrak{C} \subset H^1(\mathbb{T}^n, \mathbb{R})$  be a connected set where we are going to define  $c$ -equivalence. For each class  $c \in \mathfrak{C}$ , we assume that there exists a non-degenerate embedded  $(n-1)$ -dimensional torus  $\Sigma_c \subset \mathbb{T}^n$  such that each  $c$ -semi static curve  $\gamma$  transversally intersects  $\Sigma_c$ . Let

$$\mathbb{V}_c = \bigcap_U \{i_{U*} H_1(U, \mathbb{R}) : U \text{ is a neighborhood of } \mathcal{N}(c) \cap \Sigma_c \text{ in } \mathbb{T}^n\},$$

here  $i_U: U \rightarrow M$  denotes inclusion map.  $\mathbb{V}_c^\perp$  is defined to be the annihilator of  $\mathbb{V}_c$ , i.e. if  $c' \in H^1(\mathbb{T}^n, \mathbb{R})$ , then  $c' \in \mathbb{V}_c^\perp$  if and only if  $\langle c', h \rangle = 0$  for all  $h \in \mathbb{V}_c$ . Clearly,

$$\mathbb{V}_c^\perp = \bigcup_U \{\ker i_U^* : U \text{ is a neighborhood of } \mathcal{N}(c) \cap \Sigma_c \text{ in } \mathbb{T}^n\}.$$

Note that there exists a neighborhood  $U$  of  $\mathcal{N}(c) \cap \Sigma_c$  such that  $\mathbb{V}_c = i_{U*} H_1(U, \mathbb{R})$  and  $\mathbb{V}_c^\perp = \ker i_U^*$  (see [M93]).

**Definition 6.1** ( $c$ -equivalence). *We say that  $c, c' \in H^1(M, \mathbb{R})$  are cohomologically equivalent if there exists a continuous curve  $\Gamma: [0, 1] \rightarrow \mathfrak{C}$  such that  $\Gamma(0) = c$ ,  $\Gamma(1) = c'$ ,  $\alpha(\Gamma(s))$  keeps constant for all  $s \in [0, 1]$ , and for each  $s_0 \in [0, 1]$  there exists  $\epsilon > 0$  such that  $\Gamma(s) - \Gamma(s_0) \in \mathbb{V}_{\Gamma(s_0)}^\perp$  whenever  $s \in [0, 1]$  and  $|s - s_0| < \epsilon$ .*

**Lemma 6.1.** *For  $C^5$ -generic  $V_2$ , there exists  $\Delta_0 > 0$  as well as suitably small numbers  $\varepsilon_0 > 0$ ,  $\delta_3 \gg \delta_4 \gg \dots \gg \delta_n > 0$ , such that for each  $\varepsilon \in (0, \varepsilon_0)$ , any two points on the circle  $\Gamma_{H_{n-1}}(\hat{c}_{n-3}^*, E)$  are  $c$ -equivalent provided*

$$E - \frac{1}{2} \|\hat{c}_{n-3}^*\|_{\ell_2}^2 \in (d\delta_3, \Delta_0 - d\delta_3).$$

(see the paragraph before Corollary 6.1 for the relation between  $\hat{c}^*$  and  $\hat{c}^*$ .)

*Proof.* We extend the configuration space from  $\mathbb{T}^{n-1}$  to  $\mathbb{T}^n$ , the extra dimension is for  $x_n$ . Due to Theorem 6.2 and Corollary 6.1, we have

$$\mathcal{N}_{H_{n-1}}(\tilde{c}, \hat{c}, \alpha_{Y_\delta}(\tilde{c}, \hat{c}))|_{x_n=s} \subset \bigcup I_{\tilde{c},i} \times \left\{ (\hat{x}, x_n) \in \mathbb{T}^{n-2} \right\}$$

where  $I_{\tilde{c},i}$  is given in Proposition 6.1 with  $(\tilde{c}, \hat{c}^*) \in \tilde{\alpha}_{Y_\delta}^{-1}(E)$  and  $E - \frac{1}{2} \|\hat{c}_{n-3}^*\|_{\ell_2}^2 \in (d\delta_3, \Delta_0 - d\delta_3)$

In the configuration space  $\mathbb{T}^n$  we obtain a section  $\Sigma_{\tilde{c}} \times \mathbb{T}^{n-2}$ . Obviously, the first homology of  $\mathcal{N}_{H_{n-1}}(c) \cap (\Sigma_{\tilde{c}} \times \mathbb{T}^{n-2})$  is at most  $(n-2)$ -dimensional:

$$\mathbb{V}_c = \text{span}(e_3, e_4, \dots, e_n) \quad \text{with } c = (\tilde{c}, \hat{c}, \alpha_{Y_\delta}(\tilde{c}, \hat{c})).$$

For any two classes  $c, c' \in \Gamma_{H_{n-1}}(\hat{c}^*, E)$  with  $E - \frac{1}{2} \|\hat{c}_{n-3}^*\|_{\ell_2}^2 \in (d\delta_3, \Delta_0)$ , one can see the condition (6.4) holds, namely,  $c - c' \in \mathbb{V}_c^\perp$ . This proves our claim.  $\square$

**Remark 6.1.** Let us explain how the orbits constructed via cohomology equivalence look like in the phase space. Let us consider Hamiltonian systems of two degrees of freedom of the form  $H(x, y) = \frac{1}{2} \langle y, Ay \rangle + V(x)$ ,  $(x, y) \in T^*\mathbb{T}^2$ . On the zeroth energy level we assume that there is one unique hyperbolic fixed point, and for simplicity, that there are three edges on the boundary of flat of the  $\alpha$  function corresponding to three homoclinic orbits. In a neighborhood of the hyperbolic fixed point where no KAM tori exists, we expect to get small oscillations of the variable  $y$  in a controlled way. Indeed, consider energy levels slightly higher than zero, there are three hyperbolic periodic orbits close to the corresponding three homoclinic orbits. Moreover, the three periodic orbits all get close to the hyperbolic fixed point as the energy decreases to zero. For sufficiently low energy, the stable manifold of one periodic orbit intersects transversally the unstable manifold of another and vice versa (**H2**). In this setting, the Mañé set does not cover the whole torus so that we get a path in the  $H^1(\mathbb{T}^2, \mathbb{R})$  which is the level set of the  $\alpha$ -function along which all the cohomology classes are equivalent. In the phase space by the  $\lambda$ -lemma, we get orbits moving from one periodic orbit to another. Actually, something even stronger can be done. If we label the three periodic orbits by 1, 2, 3, then for all prescribed bi-infinite symbolic sequence in  $\{1, 2, 3\}^{\mathbb{Z}}$ , there is an orbit visiting the three periodic orbits according to the given sequence ([GT1, L]).

**6.3. Overlapping property.** We are going to show that the NHIC after all the  $n-2$  steps of reduction enters the annulus  $\mathbb{A}(\Delta_0, \hat{E})$  (see Formula (6.5)).

For the system  $\tilde{G}$  defined in (3.4), there exists an annulus  $\mathbb{A}$  of width  $\Delta_0$  surrounding the flat  $\mathbb{F}_0$ , see Corollary (6.1), which admits a foliation of  $c$ -equivalent circles. We claim that the NHIC enters the annulus region  $\mathbb{A}$  in the sense that the NHIC contains some Mather sets for those first cohomology classes lying in  $\mathbb{A}$ . Indeed, we first know the value  $\Delta_0$  for the system  $\tilde{G}$  before further reduction of order. Then, Proposition 3.1 is applied to introduce the canonical action-angle variables restricted on the cylinder all the way to the homoclinic orbit (all the way to the  $\partial\mathbb{F}$  in the  $H^1(\mathbb{T}^n, \mathbb{R})$ ). In virtue of the estimate of the derivatives of the canonical transformation in Lemma 3.4, we

choose the lower bound  $\gamma$  in Proposition 3.2 and Remark 3.2 such that the energy level  $\tilde{h} = \gamma \leq \Delta_0/2$  and fix it as the lowest allowable energy level for us to do the reduction of order. This  $\gamma$  is chosen to be independent of any of  $\delta_i$ ,  $i \geq 3$ .

As all periodic orbits in such a cylinder share the same class, e.g.  $g \in H_1(\mathbb{T}^2, \mathbb{Z})$ , and each of them support a minimal measure for certain cohomology class, all of these classes make up a channel in  $H^1(\mathbb{T}^2, \mathbb{R})$ , which is denoted by  $\mathbb{W}_g = \cup_{\lambda > 0} \mathcal{L}_{\beta_{\tilde{\mathbf{c}}}}(\lambda g)$ . Because of Lemma 5.1, it has certain width which may vary with  $\lambda > 0$ . When we turn on the  $\delta_i$ ,  $i \geq 3$  perturbations in  $\mathbf{Y}_\delta$ , the channels may become disjoint from the disc. However we have the following

**Proposition 6.2** (Overlap property). *Consider the system  $\mathbf{Y}_\delta$ . There exists a positive number  $\delta_* > 0$  such that both arms of the channel  $\mathbb{C}$  intersects the annulus-shaped region:  $\mathbb{C} \cap \mathbb{A}(\Delta_0, \tilde{E}) \neq \emptyset$  provided  $0 < \delta_3 \leq \delta_*$ .*

*Proof.* For one step of order reduction, one has Hamiltonian of the form (3.30) where  $\tilde{h}$  is well-defined provided  $\tilde{h} \geq \gamma$ . Applying the result in [CZ1] one obtains NHICs in  $\{\tilde{h} \geq \gamma\}$  if one ignores the term  $k_\delta^*(\delta R_I(\mathbf{x}) + \varepsilon^\sigma R_{II}(\mathbf{x}, \mathbf{y}))$  in (3.30). Applying Theorem of normally hyperbolic invariant manifold one obtains overflow NHICs, restricted on which the system turns out to be nearly integrable with two less degrees of freedom after one introduces new action-angle variable again. By applying a result of [BLZ] on overflow NHICs, for sufficiently small  $\delta_2 \gg \delta_4 \gg \dots \gg \delta_{n-1} \gg \varepsilon > 0$ , a three-dimensional NHIC exist which contains Aubry set for those  $c \in \mathbb{C}$  such that  $\alpha(c) \geq \gamma + \frac{1}{2}(\Delta_0 - \gamma)$  (similar to the proof of Proposition 5.1 of [C15a]).  $\square$

**6.4. Construction of the ladder.** We have completed the first step of the strategy outlined in Section 6.1. The  $\alpha_{\mathbf{Y}_\delta}$  is a  $\delta_3$  perturbation of  $\alpha_{\mathbf{Y}_0} = \tilde{\alpha}(\tilde{\mathbf{c}}) + \frac{1}{2}\|\tilde{\mathbf{c}}\|^2$ . We first move along one arm of the channel  $\mathbb{C}$  until the energy of the subsystem  $\tilde{\mathbf{G}}$  becomes in  $(d\delta_3, \Delta_0 - d\delta_3)$  at which time we fix  $\tilde{\mathbf{c}} = \tilde{\mathbf{c}}^*$  and apply the  $c$ -equivalence mechanism in the previous section to move  $\tilde{\mathbf{c}} \rightarrow -\tilde{\mathbf{c}}$ . We remark that it is enough to send  $\tilde{\mathbf{c}}$  into a small  $O_{\delta_3 \rightarrow 0}(1)$  neighborhood of  $-\mathbf{c}$  so that the image lies in  $\tilde{\mathbb{C}}$ . In this section, we work on the second step in the strategy to build the vertical curve in Figure 2 that we call ladder.

In the system  $\mathbf{Y}_\delta$ , we set  $\delta_4 = \delta_5 = \dots = \delta_n = 0$  to obtain the following system by discarding the integrable part

$$(6.6) \quad \mathbf{Y}_{3,\delta_3} = \frac{1}{2}\langle \tilde{A}\tilde{\mathbf{y}}, \tilde{\mathbf{y}} \rangle + V_2(\tilde{\mathbf{x}}) + \frac{1}{2}\mathbf{y}_3^2 + \delta_3 \mathbf{V}_3(\tilde{\mathbf{x}}_3)$$

defined on  $T^*\mathbb{T}^2 \times \mathbb{R}^2$ , where  $\tilde{\mathbf{x}} = (\mathbf{x}_1, \mathbf{x}_2)$ ,  $\tilde{\mathbf{y}} = (\mathbf{y}_1, \mathbf{y}_2)$ ,  $\tilde{\mathbf{x}}_3 = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ ,  $\tilde{\mathbf{y}}_3 = (\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3)$ . The system  $\mathbf{Y}_{3,\delta_3}$  is obtained from

$$(6.7) \quad Y_{3,\delta_3} = \frac{1}{2}\langle \tilde{A}_3\tilde{y}_3, \tilde{y}_3 \rangle + V_2(\tilde{x}) + \delta_3 V_3(\tilde{x}_3)$$

defined on  $T^*\mathbb{T}^3$  via the following linear transformation

$$\mathbf{S}_3 \tilde{x}_3 = \tilde{\mathbf{x}}_3, \quad \mathbf{S}_3^{-t} \tilde{y}_3 = \tilde{\mathbf{y}}_3,$$

where  $\mathbf{S}_3$  is the first  $3 \times 3$  block of  $\mathbf{S}$ . (See the bracketed terms in (3.41) in Section 3.10 and (4.13) with  $j = 3$  in Section 4.2.3). Next, we perform the reduction of order in the subsystem  $\tilde{\mathbf{G}} = \frac{1}{2}\langle \tilde{A}\tilde{\mathbf{y}}, \tilde{\mathbf{y}} \rangle + V_2(\tilde{\mathbf{x}})$  in  $\mathbf{Y}_{3,\delta_3}$  to get a system defined on  $T^*\mathbb{T}^1 \times \mathbb{R}^2$

$$(6.8) \quad \tilde{\mathbf{G}}_{2,\delta_3} = \tilde{h}(I) + \frac{1}{2}\mathbf{y}_3^2 + \delta_3 Z_{1,3}(I, \varphi, \mathbf{x}_3)$$

where  $Z_{1,3}$  is obtained from  $V_3$  by restricting to the NHIC.

We use the linear transformation  $\underline{\mathbf{S}}_3(\varphi, \mathbf{x}_3) := (\varphi, x_3)$ ,  $\underline{\mathbf{S}}_3^{-t}(I, \mathbf{y}_3) := (J, y_3)$  where

$$\underline{\mathbf{S}}_3 = \begin{bmatrix} 1 & 0 \\ \frac{s_3}{\sqrt{b_3}} & \frac{1}{\sqrt{b_3}} \end{bmatrix}$$

to get a system defined on  $T^*\mathbb{T}^2$

$$(6.9) \quad \tilde{G}_{2,\delta_3} = \left[ \tilde{h}\left(J + \frac{s_3}{\sqrt{b_3}}y_3\right) + \frac{1}{2}y_3^2 + \delta_3 Z_{1,3}\left(J + \frac{s_3}{\sqrt{b_3}}y_3, \varphi, \sqrt{b_3}x_3 - s_3\varphi\right) \right].$$

To see that the resulting system is defined on  $T^*\mathbb{T}^2$ , it is enough to notice that  $\tilde{G}_{2,\delta_3}$  is a rescaling of the bracketed term in (4.17), which is defined on  $T^*\mathbb{T}^2$ .

Next, let us look at the effect of the reduction on the  $\alpha$  function and the cohomology classes. Consider  $\tilde{\alpha}$  for the system  $\tilde{\mathbf{G}}$  first. It has a channel  $\tilde{\mathbf{C}}$  corresponding to the NHIC in the phase space with homology class  $(1, 0)$  (see Figure 1). The NHIC is formed by periodic orbits on different energy levels of  $\tilde{\mathbf{G}}$ . On each energy level there is only one periodic orbits except for finitely many energy levels there are two. Each hyperbolic periodic orbits corresponds to an interval in  $H^1(\mathbb{T}^2, \mathbb{R})$  according to Lemma 5.1, hence the channel is foliated into intervals lying on different energy levels of the  $\tilde{\alpha}$ . We can introduce new coordinates  $c_{\parallel}, c_{\perp}$  to reparametrize the channel  $\tilde{\mathbf{C}}$  such that  $\tilde{\alpha}(c_{\parallel}, c_{\perp}) = \tilde{\alpha}(c'_{\parallel}, c_{\perp})$  and  $\tilde{\alpha}(c_{\parallel}, c_{\perp}) \neq \tilde{\alpha}(c_{\parallel}, c'_{\perp})$  for  $c_{\perp} \neq c'_{\perp}$ . Namely  $c_{\parallel}$  parametrizes the short interval on an energy level and  $c_{\perp}$  parametrizes the energy levels.

Recall that we restrict to the NHIC in the system  $\tilde{\mathbf{G}}$  to get a system of one degree of freedom that is integrable. The reduced Hamiltonian system  $\tilde{h}(I)$  is its own  $\alpha$  function for integrable systems. This reduction procedure amounts to identify  $\tilde{\alpha}(c_{\parallel}, c_{\perp}) = \tilde{h}(I)$  and  $I = c_{\perp}$ .

The  $\alpha$  functions  $\alpha_{Y_{3,\delta_3}}, \alpha_{\tilde{G}_{2,\delta_3}}$  are defined in the standard way and we define the  $\alpha$  functions for  $\mathbf{Y}_{3,\delta_3}, \tilde{\mathbf{G}}_{2,\delta_3}$  as

$$\alpha_{\mathbf{Y}_{3,\delta_3}} = \underline{\mathbf{S}}_3^{t*} \alpha_{Y_{3,\delta_3}}, \quad \alpha_{\tilde{\mathbf{G}}_{2,\delta_3}} = \underline{\mathbf{S}}_3^{t*} \alpha_{\tilde{G}_{2,\delta_3}}.$$

The same argument as Lemma 5.2 gives that the  $\alpha$  function for  $\tilde{\mathbf{G}}_{2,0}$  is the same as the Hamiltonian

$$\alpha_{\tilde{\mathbf{G}}_{2,0}} = \tilde{h}(c_{\perp}) + \frac{1}{2}\mathbf{c}_3^2$$

and that the function  $\alpha_{\tilde{\mathbf{G}}_{2,\delta_3}}$  is a  $\delta_3$  perturbation of  $\alpha_{\tilde{\mathbf{G}}_{2,0}}$ . The function  $\alpha_{\tilde{\mathbf{G}}_{2,0}}$  has reflection symmetries

$$\alpha_{\tilde{\mathbf{G}}_{2,0}}(c_{\perp}, \mathbf{c}_3) = \alpha_{\tilde{\mathbf{G}}_{2,0}}(c_{\perp}, -\mathbf{c}_3) = \alpha_{\tilde{\mathbf{G}}_{2,0}}(-c_{\perp}, \mathbf{c}_3) = \alpha_{\tilde{\mathbf{G}}_{2,0}}(-c_{\perp}, -\mathbf{c}_3).$$

We define the first piece of ladder by (see Figure 4b)

$$\mathbb{L}_3(\mathbf{c}_3^*) := \{(c_{\perp}, \mathbf{c}_3) \mid \alpha_{\tilde{\mathbf{G}}_{2,\delta_3}}(c_{\perp}, \mathbf{c}_3) = E, \quad |\mathbf{c}_3| \leq |\mathbf{c}_3^*|, \quad c_{\perp} < 0\}$$

for the system  $\tilde{\mathbf{G}}_{2,\delta}$ , which transforms to a piece of ladder (see Figure 4a) for the system  $\tilde{G}_{2,\delta_3}$  by  $\underline{\mathbf{S}}_3^t$ . Moreover, since we have taken quotient of  $c_{\parallel}$ , in system  $\mathbf{Y}_{3,\delta_3}$  before the reduction, this ladder  $\mathbb{L}_3(\mathbf{c}_3^*)$  is embedded in a piece of two dimensional surface which is a product of  $\mathbb{L}_3(\mathbf{c}_3^*)$  with intervals. We use the same notation  $\mathbb{L}_3(\mathbf{c}_3^*)$  to denote the ladder for all the systems  $Y_{3,\delta_3}, \tilde{G}_{2,\delta_3}$  and  $\tilde{\mathbf{G}}_{2,\delta_3}$  without introducing new notations.

The next lemma shows how to move the third component of the cohomology class along the ladder.

**Lemma 6.2.** *Under generic perturbation  $\delta_3 V_3$  in the system  $Y_{3,\delta_3}$  (6.7), each Aubry set  $\tilde{\mathcal{A}}(c)$  for  $c \in \mathbb{L}_3(\mathbf{c}_3^*)$  can be connected to another Aubry set  $\tilde{\mathcal{A}}(c')$  nearby if  $c' \in \mathbb{L}_3(\mathbf{c}_3^*)$ .*

*Proof.* The system  $Y_{3,\delta_3}$  of three degrees of freedom, when restricted on its NHIM, becomes a system  $\tilde{G}_{2,\delta_3}$  of with two degrees of freedom. Both Aubry sets  $\tilde{\mathcal{A}}(c)$  and  $\tilde{\mathcal{A}}(c')$  are located in a 2-torus. We choose the covering space  $\mathbb{T} \times 2\mathbb{T} \times \mathbb{T} \ni (x_1, x_2, x_3)$ , where each Aubry set is lifted to two copies. So what we need to check is that the Mañé set does not cover the whole  $\mathbb{T} \times 2\mathbb{T} \times \mathbb{T}$ .

For the reduced Hamiltonian  $\tilde{G}_{2,\delta_3}$  restricted on an energy level set, since it has two degrees of freedom, all elementary weak KAMs can be parameterized by certain “volume”  $\sigma$  so that this family of weak KAMs is  $\frac{1}{3}$ -Hölder continuous in  $\sigma \in I$  (see Theorem 9.2), where  $I$  is a closed interval. We choose a subset  $\Delta \subset I$  in the following way:  $\sigma \in I$  if and only if the weak KAM  $u_\sigma$  is  $C^1$  (must be  $C^{1,1}$  also [FS, Be4]), i.e. the Mañé set is an invariant 2-torus. For  $\sigma \notin \Delta$ , certain section  $\Sigma_\sigma$  of 2-torus exists such that  $\mathcal{N}_\sigma \cap \Gamma_\sigma$  is shrinkable.

Since the reduced system  $\tilde{G}_{2,\delta_3}$  lies on a normally hyperbolic manifold, we find that the barrier function  $B_\sigma$  of the system  $Y_{3,\delta_3}$  is  $\frac{1}{6}$ -Hölder continuous in  $\sigma$  by applying Theorem 9.3. This property allows us to show the following holds for generic  $\delta_3 V_3$  (perturbing  $V_3$  at the place away from the NHIM)

*for all  $\sigma \in \Delta$ , each connected component of  $\text{Argmin}\{B_\sigma, \Sigma_{0,\sigma} \setminus \cup_m N_m\}$  is contained in certain disk  $O_m \subset \Sigma_{0,\sigma}$ ,*

where  $\Sigma_{0,\sigma}$  is a 2-dimensional section of  $\mathbb{T}^3 \ni (x_1, x_2, x_3)$  which is transversal to  $c(\sigma)$ -semi static curves,  $\cup_m N_m$  denotes a neighborhood of the Aubry set in the finite covering space,  $\text{Argmin}\{B_\sigma, \Sigma_0 \setminus \cup_m N_m\}$  denotes the set of minimal points of  $B_\sigma$  which fall into the set  $\Sigma_{0,\sigma} \setminus \cup_m N_m$ .

The proof of this statement is the same as in the proof in Section 8.3 of [C12]. We can not apply the argument in [CY1, CY2], as the perturbation  $\delta_3 V_3$  is only allowed to depend on  $x$ .

Therefore, the Aubry set can be connected to another one nearby, either by connecting orbit of type- $h$  when the Mañé set is a 2-torus (Arnold’s mechanism of intersection of (un)stable “manifolds”), or by type- $c$  when the Mañé set does not cover the whole 2-torus (Mather’s mechanism of crossing the Birkhoff instability region) (For the definition of local connecting orbit of type- $h$  with incomplete intersection as well as type- $c$ , refer to Section 7.1).  $\square$

In the next remark, we explain our mechanism of ladder climbing.

**Remark 6.2** (The diffusion mechanism for the ladder climbing). Here we employ a variant of Arnold’s mechanism (1.2). Consider Hamiltonian system of three degrees of freedom of the form

$$H = \frac{y_1^2}{2} + \frac{y_2^2}{2} + \frac{y_3^2}{2} + (\cos x_3 - 1)(1 + \varepsilon(\cos x_1 + \sin x_2)).$$

In this system, there exists diffusing orbit for each  $E > 0$  such that  $(y_1, y_2)$  stays close to the circle  $\{y_1^2 + y_2^2 = 2E\}$  and  $\arctan \frac{y_1}{y_2}$  achieves any value in  $[0, 2\pi)$ . Loosely speaking,  $(y_1, y_2)$  moves along the circle  $\{y_1^2 + y_2^2 = 2E\}$ . In the example, the set

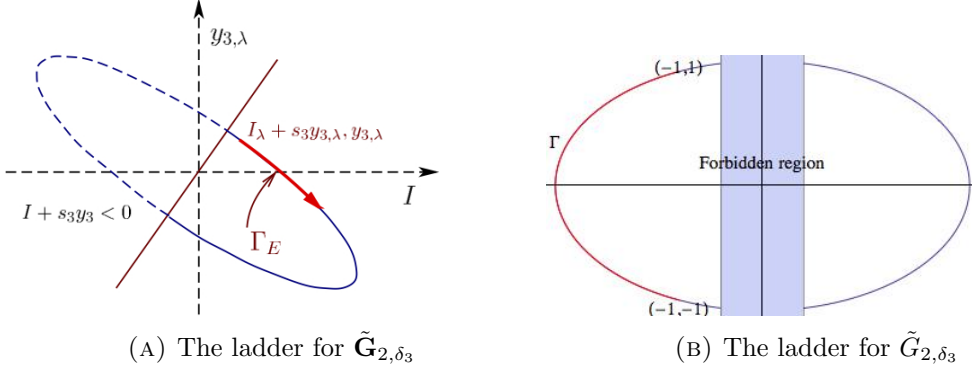


FIGURE 4. The ladder construction

$\{x_3 = y_3 = 0\}$  is a normally hyperbolic invariant manifold and the  $\varepsilon$  perturbation vanishes on the manifold as in Arnold's example (1.2). This can be considered as a system of *a priori* unstable type. One can compute the Melnikov integral as in the Arnold's example to verify that Arnold diffusion exists. In our case, the system  $Y_{3,\delta_3}$  plays the role of  $H$  here and the system  $\tilde{G}_{2,\delta_3}$  plays the role of  $y_1^2 + y_2^2$ .

**Remark 6.3.** Since we have required  $|I| > \gamma$  for some  $\gamma < \Delta_0/2$  to avoid the singular behavior of the coordinates  $(I, \varphi)$  estimated in Lemma 3.4. Due to the convexity and symmetricity of the  $\alpha$  function, the ladder never touches the forbidden region  $|I| \leq \gamma$  (see Figure 4b).

Lemma 6.2 enables us to move  $\mathbf{c}_3$  from  $\mathbf{c}_3^*$  to  $-\mathbf{c}_3^*$ . In the next theorem, we apply Lemma 6.2 repeatedly combined with our reduction of order scheme to build ladders to move all the components of  $\hat{\mathbf{c}}_{n-3}^* \rightarrow -\hat{\mathbf{c}}_{n-3}^*$ .

**Theorem 6.3.** *For the Hamiltonian of  $Y_\delta$  (4.19) with generic  $V_2, V_3, \dots, V_{n-1}$  and small  $E > 0$ , there exists  $0 < \delta_{n-1} \ll \dots \ll \delta_4 \ll \delta_3$  such that following holds:*

*Let  $\mathbb{C}, \mathbb{C}'$  be two channels connected to the flat  $\mathbb{F}_0$  defined in (5.9). For a class  $(\tilde{c}, \hat{c}) \in \mathbb{C} \cap \alpha_{Y_\delta}^{-1}(E)$ , there is  $(\tilde{c}', \hat{c}') \in \mathbb{C}' \cap \alpha_{Y_\delta}^{-1}(E)$  which is connected to a class  $(\tilde{c}^*, \hat{c})$  by a ladder  $\mathbb{L}$ . For any two classes  $c^*, c'' \in \mathbb{L}$  with small  $|c^* - c''|$ , the Aubry sets  $\tilde{\mathcal{A}}(c^*)$  and  $\tilde{\mathcal{A}}(c'')$  are connected by local minimal orbit either of type-h with incomplete intersection or of type-c.*

*Proof.* We again work with the system  $\mathbf{Y}_\delta$  then transform the obtained ladder to the system  $Y_\delta$  using the linear transformation induced by  $\mathbf{S}$ . Applying Lemma 6.2, we get a ladder  $\mathbb{L}_3$  connecting two points  $(-c_\perp, \mathbf{c}_3^*)$  and  $(-c_\perp, \mathbf{c}_3^*)$  in system  $\tilde{\mathbf{G}}_{2,\delta_3}$ . Keep it in mind that such a ladder is constructed for fixed  $\hat{\mathbf{c}}_{n-4}^*$ .

For the construction of the ladder moving  $\mathbf{c}_4$  from  $\mathbf{c}_4^*$  to  $-\mathbf{c}_4^*$ , we recover the potential perturbation  $\delta_4 V_4$ . For generic  $V_4$ , the reduced Hamiltonian can be further reduced to the following (see (4.21) with  $i = 2$ )

$$\begin{aligned}
 \bar{Y}_{2,\delta} = & \left[ \tilde{h}_3(J_3 + s_4 y_4) + \frac{1}{2} b_4 y_4^2 + \delta_4 Z_{1,4}(J_3 + s_4 y_4, \varphi_3, x_4 - s_4 \varphi_3) \right] \\
 (6.10) \quad & + \sum_{k=5}^{n-1} \left( \frac{b_k}{2} y_k^2 + \delta_k Z_{1,k}(J_3, \varphi_3, x_4, x_5, \dots, x_k) \right) + \delta_n \bar{R}_2.
 \end{aligned}$$

where the bracketed term is the restriction to the NHIC in the system defined on  $T^*\mathbb{T}^4$  (see the bracketed term in (4.13) with  $j = 4$ )

$$Y_{4,\delta_3,\delta_4} = \frac{1}{2} \langle \tilde{A}_4 \tilde{y}_4, \tilde{y}_4 \rangle + V_2(\tilde{x}) + \delta_3 V_3(\tilde{x}_3) + \delta_3 V_4(\tilde{x}_4).$$

By applying Lemma 6.2 again, we obtain a ladder  $\mathbb{L}_4$  connecting two cohomology classes  $(c_\perp, -\mathbf{c}_3^*, \mathbf{c}_4^*, \mathbf{c}_5^*, \dots, \mathbf{c}_{n-1}^*)$  and  $(c_\perp, -\mathbf{c}_3^*, -\mathbf{c}_4^*, \mathbf{c}_5^*, \dots, \mathbf{c}_{n-1}^*)$ .

By induction combined with (4.21), we obtain  $n - 3$  ladders  $\mathbb{L}_3, \dots, \mathbb{L}_{n-1}$ , we call them *simple ladder*. The composition of these simple ladders  $\mathbb{L} = \mathbb{L}_{n-1} * \dots * \mathbb{L}_3$  connecting two points  $(-\tilde{\mathbf{c}}, \hat{\mathbf{c}}^*)$  and  $(-\tilde{\mathbf{c}}, -\hat{\mathbf{c}}^*)$ . We remark that it is enough to send the cohomology classes into a neighborhood in  $\mathbb{C}$  of  $(-\tilde{\mathbf{c}}, -\hat{\mathbf{c}}^*)$ .

To complete the argument for the construction of the ladders, we are required to show the ladder  $\mathbb{L}_\ell$  survives perturbation  $\sum_{j=\ell+1}^{n-1} \delta_j V_j$  for each  $j = 3, \dots, n - 2$ . We use Theorem 7.2 to prove it.

Because of upper-semi continuity of Mañé set (see the proof of Corollary 6.1 for how to apply it to  $Y_\delta$ ), in the finite covering space  $\mathbb{T} \times 2\mathbb{T} \times \mathbb{T}^{n-3}$ , the following condition holds for each  $c \in \mathbb{L}_3$  provided  $\delta_j \ll \delta_3$  for each  $j \in \{4, \dots, n - 1\}$ : each connected component of  $\text{Argmin}\{B_\sigma, (\Sigma_{0,\sigma} \setminus \cup_m N_m) \times \mathbb{T}^{n-4}\}$  is contained in certain  $O_m \times \mathbb{T}^{n-4}$ , where  $O_m \subset \Sigma_{0,\sigma}$  is a 2-dimensional disk. The conditions of Theorem 7.2 are satisfied. So,  $\mathbb{L}_3$  survives perturbation. For other  $\mathbb{L}_j$  with  $j \in \{4, \dots, n - 2\}$ , the proof is the same provided  $\delta_{j+1} \ll \delta_j$ .  $\square$

In this way we have established a transition chain to cross the  $m$ -strong resonance. We will show in Theorem 7.2 in Section 7 that the conclusion of Theorem 6.3 implies the existence of diffusing orbit with cohomology classes changing along  $\Gamma_\delta$ .

In the next remark, we explain the frequency space dynamics.

**Remark 6.4.** Let us keep track of the frequency vector. As  $a$  varies, let us consider the frequency line  $\omega_a^\# = \lambda_a^\#(a, \frac{p_2}{q_2}, \dots, \frac{p_n}{q_n})$ . We set  $\lambda_a^\# = 1$  for simplicity and suppose  $a = a''$  corresponds to a complete resonance. For  $a - a'' = \lambda$ , we have  $\omega_a^\# = \omega_{a''}^\# + \lambda \mathbf{e}_1$ . To cross the complete resonance amounts to move the frequency such that  $\lambda$  crosses zero. We are unable to move  $\lambda$  from positive to zero to negative or vice versa directly. Instead, we take a detour as follows. We first move  $\omega_a$  to some point where  $\lambda$  is sufficiently small. We next look at the system  $\mathbf{Y}_\delta$  in which the frequency becomes  $\mathbf{S}\lambda \mathbf{e}_1 = \lambda(1, 0, \frac{s_3}{\sqrt{b_3}}, \dots, \frac{s_n}{\sqrt{b_n}})$ . The c-equivalence mechanism amounts to send the first two components  $(\lambda, 0) \rightarrow (-\lambda, 0)$  via a path similar to  $\lambda(\cos \theta, \sin \theta)$ ,  $\theta \in [0, \pi]$ . The resulting frequency vector for the original system is now

$$\begin{aligned} & \omega_{a''}^\# + \lambda \mathbf{S}^{-1} \left( -1, 0, \frac{s_3}{\sqrt{b_3}}, \dots, \frac{s_n}{\sqrt{b_n}} \right) \\ &= \omega_{a''}^\# - \lambda \mathbf{S}^{-1} \left( 1, 0, \frac{s_3}{\sqrt{b_3}}, \dots, \frac{s_n}{\sqrt{b_n}} \right) + 2\lambda \mathbf{S}^{-1} \left( 0, 0, \frac{s_3}{\sqrt{b_3}}, \dots, \frac{s_n}{\sqrt{b_n}} \right) \\ &= \omega_{a''-\lambda}^\# + 2\lambda \mathbf{S}^{-1} \left( 0, 0, \frac{s_3}{\sqrt{b_3}}, \dots, \frac{s_n}{\sqrt{b_n}} \right) \neq \omega_{a''-\lambda}^\#. \end{aligned}$$

This error  $2\lambda \mathbf{S}^{-1} \left( 0, 0, \frac{s_3}{\sqrt{b_3}}, \dots, \frac{s_n}{\sqrt{b_n}} \right)$  in frequency corresponds to the vertical misalignment of cohomology classes in Figure 2. Our ladder are designed to kill this error step by step.

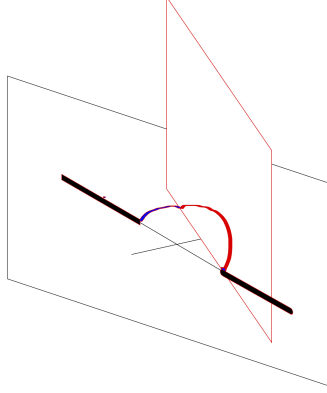


FIGURE 5. The frequency space picture corresponding to Fig 2. The black lines  $\omega_a$  lies in the plane  $\mathbf{k}'^\perp$  (the black plane). The line intersect another line  $(\mathbf{k}'^\perp \cap \mathbf{k}''^\perp)$  at a strong double resonance. The red plane is  $\text{span}\{\mathbf{k}', \mathbf{k}''\}$  since around the strong double resonance, the leading term of the potential is  $V(\langle \mathbf{k}', x \rangle, \langle \mathbf{k}'', x \rangle)$ . The red path lies in the red plane (c-equivalence mechanism). The blue path (the ladder) lies in  $\mathbf{k}'^\perp$  hence has normal hyperbolicity due to the resonance  $\mathbf{k}'$ .

## 7. VARIATIONAL CONSTRUCTION OF GLOBAL DIFFUSION ORBITS

Global diffusion orbits are constructed shadowing a sequence of local connecting orbits end to end. There are two types of local connecting orbit, one is called type- $h$  as which looks like a “heteroclinic” orbit, another one is called type- $c$  as it is constructed by using “cohomology equivalence”.

**7.1. Local connecting orbit of type- $h$  with incomplete intersections.** For an Aubry set, if its stable set “intersects” its unstable set transversally, this Aubry set is connected to any other Aubry set nearby by local minimal orbits. It can be thought as a variational version of Arnold’s mechanism, the condition of geometric transversality is replaced by the total disconnectedness of minimal points of the barrier function.

However, this condition is not always satisfied for the problem we encountered here. The stable set may intersect the unstable set on a set with nontrivial first homology, i.e. *incomplete intersection*. In this section, we design a new method to handle this problem. Let us first formulate a version for time-periodic dependent Lagrangian.

Recall the definition of the function  $h_c^\infty$  introduced in [M93]

$$h_c^\infty(x, x') = \liminf_{k \rightarrow \infty} \inf_{\substack{\gamma(-k)=x \\ \gamma(k)=x'}} \int_{-k}^k \left( L(\gamma(t), \dot{\gamma}(t), t) - \langle c, \dot{\gamma} \rangle + \alpha(c) \right) dt.$$

This function is closely related to weak KAM. Indeed, for  $x \in \mathcal{A}_{c,i}|_{t=0}$  (the time-1-section of the Aubry class  $\mathcal{A}_{c,i} \subset \mathcal{A}(c)$ ) we have

$$h_c^\infty(x, x') = u_{c,i}^-(x') - u_{c,i}^+(x),$$

where both  $u_{c,i}^-$  and  $u_{c,i}^+$  are the time-1-section of backward and forward elementary weak KAM respectively (see the Appendix A.3 for details). It inspired us to introduce a barrier function for two Aubry classes  $\mathcal{A}_{c,i}$  and  $\mathcal{A}_{c,j}$

$$B_{c,i,j}(x) = u_{c,j}^-(x) - u_{c,i}^+(x).$$



Passing through its minimal point there is a semi-static curve connecting these two classes, provided this point does not lie in the Aubry set.

If the Aubry set contains only one class, we work in certain finite covering space so that there are two classes. For example, if the configuration space is  $\mathbb{T}^{j+k+\ell}$  and the time-1-section of the Aubry set stays in a neighbourhood of certain lower dimensional torus,  $\mathcal{A}_0(c) \subset \mathbb{T}^{j+\ell} + \delta$ , we introduce a covering space  $\mathbb{T}^{j+\ell} \times \mathbb{T}^{k-1} \times 2\mathbb{T}$ . With respect to this covering space the Aubry set contains two classes.

We introduce some notation and conventions. For the product space  $\mathbb{T}^{j+k+\ell}$  we use  $\mathbb{T}^{j+\ell} = \{x \in \mathbb{T}^{j+k+\ell} : x_i = 0 \ \forall \ i = j+1, \dots, j+k\}$ . Given a set  $S$ , a point  $x$  and a number  $\delta$ ,  $S + x$  denotes the translation of  $S$  by  $x$ , i.e.  $S + x = \{x' + x : x' \in S\}$  and  $S + \delta$  denotes  $\delta$ -neighborhood of  $S$ , i.e.  $S + \delta = \{x : d(x, S) \leq \delta\}$ . A set  $N$  is called neighborhood of  $(j, \ell)$ -torus if it is homeomorphic to an open neighborhood of  $(j + \ell)$ -dimensional torus whose first homology group is generated by  $\{e_i : i = 1, \dots, j, j+k+1, \dots, j+k+\ell\}$ . Given a function  $B$ , we use  $\text{Argmin}\{B, S\} = \{x \in S : B(x) = \min B\}$  to denote the set of those minimal points of  $B$  which are contained in the set  $S$ .

**Theorem 7.1.** *For a time-periodic  $C^2$ -Lagrangian  $L : T\mathbb{T}^{j+k+\ell} \times \mathbb{T} \rightarrow \mathbb{R}$  and a first cohomology class  $c \in H^1(\mathbb{T}^{j+k+\ell}, \mathbb{R})$  we assume the conditions as follows:*

- (1) *the Aubry set  $\mathcal{A}(c)$  contains two classes  $\{\mathcal{A}_{c,i}, \mathcal{A}_{c,i'}\}$  which lie in a neighbourhood of  $(j, \ell)$  torus  $\mathcal{A}_{c,i}|_{t=0} \subset N_i$  and  $\mathcal{A}_{c,i'}|_{t=0} \subset N_{i'}$ . These neighborhoods are separated, i.e.  $N_i \cap N_{i'} = \emptyset$ ;*
- (2) *there exist topological balls  $\{O_m \subset \mathbb{T}^{j+k}\}$  with  $\bar{O}_m \cap \bar{O}_{m'} = \emptyset$  for  $m \neq m'$ , each connected component of*

$$\text{Argmin}\{B_{c,i,i'}, \mathbb{T}^{j+k+\ell} \setminus N_i \cup N_{i'}\}$$

*is contained in certain  $O_m \times \mathbb{T}^\ell$ ;*

*Then, for  $c' \in H^1(\mathbb{T}^{j+k+\ell}, \mathbb{R})$  satisfying following conditions*

- (1)  *$\langle c' - c, g \rangle = 0$  holds  $\forall \ g \in H_1(\mathbb{T}^{j+k+\ell}, \mathbb{T}^{j+k}, \mathbb{Z})$  and  $|c' - c| \ll 1$ ;*
- (2) *the Aubry set  $\mathcal{A}(c') \subset N_i \cup N_{i'}$ ;*

*there exists an orbit  $(\gamma, \dot{\gamma})$  of  $\phi_L^t$  which connects  $\tilde{\mathcal{A}}(c)$  to  $\tilde{\mathcal{A}}(c')$  in the following sense, the  $\alpha$ -limit set of  $(\gamma, \dot{\gamma})$  is contained in  $\tilde{\mathcal{A}}(c)$ , the  $\omega$ -limit set of  $(\gamma, \dot{\gamma})$  is contained in  $\tilde{\mathcal{A}}(c')$  or vice versa.*

**Remark 7.1.** If  $\ell = 0$ , the set  $\text{Argmin}\{B_{c,i,i'}, \mathbb{T}^{j+k+\ell} \setminus N_i \cup N_{i'}\}$  is topologically trivial, it implies the stable set “intersects” the unstable set completely. Therefore, it turns out to be a variational version of Arnold’s mechanism.

**Remark 7.2.** If the Aubry set consists of one Aubry class, we study this problem in certain covering space so that the Aubry set consists of two classes. The second condition for  $c$  can be weakened so that the result becomes sharper, but the condition here is easier to verify and good enough for our purpose. Because of the upper semi-continuity of Mañé set in the first cohomology class, the Aubry set  $\mathcal{A}(c')$  is also contained in neighborhoods of these lower dimensional tori.

*Proof.* It is proved by exploiting the upper semi-continuity of Mañé set with respect to perturbation on the Lagrangian. As  $\mathcal{A}(c') \subset N_i \cup N_{i'}$ , without lose of generality we assume  $\mathcal{A}(c') \cap N_{i'} \neq \emptyset$ .

Given a ball  $O_m$  there exists small  $\epsilon$  such that  $O_m + \epsilon$  does not touch other balls. Let  $\tau_1: \mathbb{R} \rightarrow [0, \epsilon]$  be a smooth function such that  $\tau_1(t) = 0$  for  $t \in (-\infty, 0] \cup [1, \infty)$ ,  $\tau_1(t) \geq 0$  for  $t \in [0, 1]$  and  $\max \tau_1 = 1$ . Let  $\tau_2: \mathbb{R} \rightarrow [0, 1]$  be a smooth function such that  $\tau_2(t) = 0$  for  $t \leq 0$  and  $\tau_2(t) = 1$  for  $t \geq 1$ . Let  $v: \mathbb{T}^{j+k+\ell} \rightarrow [0, \epsilon]$  so that  $v(x) = 0$  if  $x \notin (O_m + \epsilon) \times \mathbb{T}^\ell$  and  $v(x) = \epsilon$  if  $x \in O_m \times \mathbb{T}^\ell$ . As  $\langle c' - c, g \rangle = 0$  for each  $g \in H_1(\mathbb{T}^{j+k+\ell}, \mathbb{T}^{j+k}, \mathbb{Z})$ ,  $\exists$  smooth function  $u \in \mathbb{T}^{j+k+\ell} \rightarrow \mathbb{R}$  so that  $\partial u = c' - c$  when it is restricted in  $(O_m + \epsilon) \times \mathbb{T}^\ell$  and  $\partial u = 0$  if  $x \notin (O_m + 2\epsilon) \times \mathbb{T}^\ell$ .

We introduce a modified Lagrangian

$$L_{c,v,u}(\dot{x}, x, t) = L(\dot{x}, x, t) - \langle c, \dot{x} \rangle - \tau_1(t)v(x) - \tau_2(t)\langle c' - c - \partial u, \dot{x} \rangle$$

and consider the minimizer  $\gamma_{k^-, k^+}: [-k^-, k^+] \rightarrow \bar{M}$  of the action

$$h_{c,v,u}^{k^-, k^+}(x^-, x^+) = \inf_{\substack{\gamma(-k^-)=x^- \\ \gamma(k^+)=x^+}} \int_{-k^-}^{k^+} L_{c,v,u}(\gamma(t), \dot{\gamma}(t), t) dt + k^- \alpha(c) + k^+ \alpha(c')$$

where  $x \in \mathcal{A}_{c,i}|_{t=0}$  and  $x' \in \mathcal{A}_{c',i'}|_{t=0}$ . As the Lagrangian is Tonelli, for any large  $T$ , the set of the curves  $\{\gamma_k|_{[-T, T]} : k^-, k^+ \geq T\}$  is  $C^2$ -bounded, therefore it is  $C^1$ -compact. Let  $T \rightarrow \infty$ , by diagonal extraction argument, we can find a subsequence of  $\gamma_{k_i}$  which converges  $C^1$ -uniformly on each compact interval to a  $C^1$ -curve  $\gamma: \mathbb{R} \rightarrow \bar{M}$  which is a minimizer of  $L_{c,v,u}$  on any compact interval of  $\mathbb{R}$ .

Let  $\mathcal{C}(L_{c,v,u})$  denote the set of minimal curves of  $L_{c,v,u}$ , it follows from the above argument that the set  $\mathcal{C}(L_{c,v,u})$  is non-empty. Restricted on  $(-\infty, 0]$  as well as on  $[1, \infty)$ , each curve in  $\mathcal{C}(L_{c,v,u})$  solves the Euler-Lagrange equation for  $L$  since  $\tau_1 = 0$  and  $\langle c' - c - \partial u, \dot{x} \rangle$  is closed. We are going to show that it also solves the equation for  $t \in [0, 1]$ .

If both  $\tau_1$  and  $\tau_2$  vanish, each curve in the set  $\mathcal{C}(L_{c,v,u})$  is nothing else but a  $c$ -semi static curve of  $L$ . These curves produce orbits which connect  $\mathcal{A}_{c,i}$  to  $\mathcal{A}_{c,i'}$ . Consider all semi-static curves which intersect  $O_m \times \mathbb{T}^\ell$  at  $t = 0$ . As  $O_m \times \mathbb{T}^\ell$  is open, the set of semi-static curves is closed,  $\exists$  small  $t_\delta > 0$  such that these curve intersect  $O_m \times \mathbb{T}^\ell$  also for  $t \in [0, t_\delta]$ . If we set  $\tau_1 = 0$  for  $t \in (-\infty, 0] \cup [t_\delta, \infty)$  and set  $\tau_2 \equiv 0$ , these semi-static curves solve the Euler-Lagrange equation produced by  $L_{c,v,u}$ . As a matter of fact, along these curves the function  $v$  keeps constant when  $\tau_1 \neq 0$ , the term  $\tau_1 v$  does not contribute to the equation. Clearly, the action of  $L_{c,v,u}$  along these curves is smaller than those semi-static curves which do not pass through  $O_m \times \mathbb{T}^\ell$  around  $t = 0$ . Since  $L_{c,v,u}$  is no longer time-periodic, a time-1-translation of its minimal curve is not necessarily minimal, i.e.  $\gamma \in \mathcal{C}(L_{c,v,u})$  does not guarantee  $k^* \gamma \in \mathcal{C}(L_{c,v,u})$  for  $k \in \mathbb{Z}$ , where  $k^*$  denotes a translation operator such that  $k^* \tau(t) = \tau(t + k)$ .

Next, let us recover the term  $\tau_2$ . Because of upper semi-continuity, the minimal curve of  $L_{c,v,u}$  must pass through  $O_m \times \mathbb{T}^\ell$  if  $c'$  is sufficiently close to  $c$ . Again, along these curves, the term  $\tau_2 \partial u$  does not contribute to the Euler-Lagrange equation, along these curves  $\partial u = c' - c$  when  $\tau_2 \in (0, 1)$ .

Obviously, the orbit produced by each curve in the set  $\mathcal{C}(L_{c,v,u})$  takes  $\tilde{\mathcal{A}}(c)$  as its  $\alpha$ -limit set and take  $\tilde{\mathcal{A}}(c')$  its  $\omega$ -limit set.  $\square$

The orbit  $(\gamma, \dot{\gamma})$  obtained in this theorem is locally minimal in the following sense:

**Local minimum:** *There are open balls  $V_i^-$ ,  $V_{i'}^+$  and positive integers  $t^-, t^+$  such that  $\bar{V}_i^- \subset N_i \setminus \mathcal{A}_0(c)$ ,  $\bar{V}_{i'}^+ \subset N_{i'} \setminus \mathcal{A}_0(c')$ ,  $\gamma(-k^-) \in V_i^-$ ,  $\gamma(k^+) \in V_{i'}^+$  and*

$$(7.1) \quad \begin{aligned} & h_c^\infty(x^-, m_0) + h_{c,v,u}^{k^-, k^+}(m_0, m_1) + h_{c'}^\infty(m_1, x^+) \\ & - \liminf_{\substack{k_i^- \rightarrow \infty \\ k_i^+ \rightarrow \infty}} \int_{-k_i^-}^{k_i^+} L_{c,v,u}(d\gamma(t), t) dt - k_i^- \alpha(c) - k_i^+ \alpha(c') > 0 \end{aligned}$$

*holds  $\forall (m_0, m_1) \in \partial(V_i^- \times V_{i'}^+)$ ,  $x^- \in N_i \cap \pi_x(\alpha(d\gamma))_{t=0}$ ,  $x^+ \in N_{i'} \cap \pi_x(\omega(d\gamma))_{t=0}$ , where  $k_i^-, k_i^+ \in \mathbb{Z}^+$  are the sequences such that  $\gamma(-k_i^-) \rightarrow x^-$  and  $\gamma(k_i^+) \rightarrow x^+$ .*

The set of curves starting from  $V_i^-$  and reaching  $V_{i'}^+$  with time  $k^- + k^+$  make up a neighborhood of the curve  $\gamma$  in the space of curves. If it touches the boundary of this neighborhood, the action of  $L_{c,v,u}$  along a curve  $\xi$  will be larger than the action along  $\gamma$ . The local minimality is crucial in the variational construction of global connecting orbits.

Next, we formulate the theorem for autonomous Lagrangian. As the Lagrangian is independent of time, one angle variable plays the role of time. Given a first cohomology class, some coordinate system exists  $G_c^{-1}x$  such that  $\omega_1(\mu) > 0$  for each ergodic  $c$ -minimal measure  $\mu$  if  $\alpha(c) > \min \alpha$ , where we use  $\omega(\mu) = (\omega_1(\mu), \dots, \omega_n(\mu))$  to denote the rotation vector of the invariant measure (see [Lx]). For this purpose, we work in a covering space  $\bar{\pi} : \bar{M} = \mathbb{R} \times \pi_{-1}\check{M}$ , where  $\pi_{-1}$  denotes the operation to eliminate the first entry,  $\pi_{-1}(x_1, x_2, \dots, x_m) = (x_2, \dots, x_m)$ , the dimension  $\mathbb{R}$  is for the coordinate  $x_1$ ,  $\check{M} = \mathbb{T}^{j+\ell} \times \mathbb{T}^{k-1} \times 2\mathbb{T}$  if the Aubry set consists of only one class which stays in a neighbourhood of  $(j, \ell)$ -torus and  $\check{M} = \mathbb{T}^{j+k+\ell}$  if the Aubry set contains two classes.

**Theorem 7.2.** *For the autonomous  $C^2$ -Lagrangian  $L : T\mathbb{T}^{j+k+\ell} \rightarrow \mathbb{R}$  and the first cohomology class  $c \in H^1(\mathbb{T}^{j+k+\ell}, \mathbb{R})$  we assume the conditions as follows:*

- (1)  $\omega_1(\mu) > 0$  holds for each ergodic  $c$ -minimal measure.
- (2) the Aubry set  $\mathcal{A}(c, \check{M})$  contains two classes  $\{\mathcal{A}_{c,i}, \mathcal{A}_{c,i'}\}$ , both stay in a neighbourhood of  $(j, \ell)$  torus, i.e.  $\mathcal{A}_{c,i} \subset N_i$ ,  $\mathcal{A}_{c,i'} \subset N_{i'}$ . These neighborhoods are separated, i.e.  $\bar{N}_i \cap \bar{N}_{i'} = \emptyset$ . The lift of both  $N_i$  and  $N_{i'}$  to  $\bar{M}$  is still connected and extends to  $x_1 = \pm\infty$ ;
- (3) there exist topological disks  $\{O_m \subset \pi_{-1}(\mathbb{T}^j \times \mathbb{T}^k)\}$  with  $\bar{O}_m \cap \bar{O}_{m'} = \emptyset$  for  $m \neq m'$ , such that each connected component of

$$\text{Argmin}\{B_{c,i,i'}, \Sigma_0 \setminus N_i \cup N_{i'}\}$$

*is contained in certain  $\{x_1 = 0\} \times O_m \times \mathbb{T}^\ell$ , where  $\Sigma_0 = \{x_1 = 0\} \times \pi_{-1}\check{M}$  is a section of  $\bar{M}$ .*

*Then, for  $c' \in H^1(\mathbb{T}^{j+k+\ell}, \mathbb{R})$  satisfying following conditions*

- (1)  $\alpha(c') = \alpha(c)$ ;
- (2)  $\langle c' - c, g \rangle = 0$  holds  $\forall g \in H_1(\mathbb{T}^{j+k+\ell}, \mathbb{T}^{j+k}, \mathbb{Z})$  and  $|c' - c| \ll 1$ ;
- (3) the Aubry set  $\mathcal{A}(c') \subset N_i \cup N_{i'}$ ;

*there exists an orbit  $(\gamma, \dot{\gamma})$  of  $\phi_L^t$  which connects  $\tilde{\mathcal{A}}(c)$  to  $\tilde{\mathcal{A}}(c')$  in the following sense, the  $\alpha$ -limit set of  $(\gamma, \dot{\gamma})$  is contained in  $\tilde{\mathcal{A}}(c)$ , the  $\omega$ -limit set of  $(\gamma, \dot{\gamma})$  is contained in  $\tilde{\mathcal{A}}(c')$  or vice versa.*

**Remark 7.3.** For autonomous system, barrier function keeps constant along minimal curve. The intersection of minimal curves of autonomous system with the section  $\Sigma_c$  is an analogy of  $\mathcal{A}_0(c)$  and  $\mathcal{N}_0(c)$  for time-periodic system.

To prove this theorem and establish an analogous inequality of (7.1), we need some notations and definitions. A Lagrangian  $L: T\bar{M} \rightarrow \mathbb{R}$  is called space-step if there exist Lagrangian  $L^-, L^+ \in C^2(T\mathbb{T}^{j+k+\ell}, \mathbb{R})$ , such that  $L^-(x_1, \cdot)|_{(-\infty, -\delta)} = L(x_1, \cdot)|_{(-\infty, -\delta)}$  and  $L^+(x_1, \cdot)|_{(\delta, \infty)} = L(x_1, \cdot)|_{(\delta, \infty)}$  where we treat  $L^\pm: TM \rightarrow \mathbb{R}$  as its natural lift to  $T\mathbb{T}^{j+k+\ell}$ . We assume some conditions:

- (1)  $\omega_1(\mu^\pm) > 0$  for each ergodic minimal measure  $\mu^\pm$  of  $L^\pm$  respectively;
- (2)  $\min \beta_{L^-} = \min \beta_{L^+}$ , without losing of generality, it equals zero;
- (3)  $|L^- - L^+| \leq \frac{1}{2} \min_{\omega_1=0} \{\beta_{L^-}(\omega'), \beta_{L^+}(\omega')\}$ .

As the minimal average action of  $L^\pm$  is achieved on  $\text{supp} \mu^\pm$  with  $\omega(\mu^\pm) = 0$ , one has  $\min_{\omega_1(\nu) \neq 0} \int L^\pm d\nu > \min \int L^\pm d\nu$ , so the third condition makes sense. To introduce the concept of minimal curve for space-step Lagrangian, we define

$$h_L^T(\bar{m}_0, \bar{m}_1) = \inf_{\substack{\bar{\gamma}(-T) = \bar{m}_0 \\ \bar{\gamma}(T) = \bar{m}_1}} A_L(\bar{\gamma}|_{[-T, T]}), \quad \forall \bar{m}_0, \bar{m}_1 \in \bar{M},$$

where

$$A_L(\bar{\gamma}|_{[-T, T]}) = \int_{-T}^T L(\bar{\gamma}(t), \dot{\bar{\gamma}}(t)) dt.$$

To generalize semi-static curve to space-step Lagrangian, we first define a set  $\mathcal{G}(L)$  of minimal curves. Because of the following lemma (Lemma 6.2 of [C12]), one can find the proof there)

**Lemma 7.1.** *If the rotation vector of each ergodic minimal measure has positive first component  $\omega_1(\mu^\pm) > 0$ ,  $\bar{m}_0 \neq \bar{m}_1$ , then*

$$\lim_{T \rightarrow 0} h_L^T(\bar{m}_0, \bar{m}_1) = \infty \quad \text{and} \quad \lim_{T \rightarrow \infty} h_L^T(\bar{m}_0, \bar{m}_1) = \infty.$$

Consequently, the following definition makes sense

**Definition 7.1.** *A curve  $\bar{\gamma}: \mathbb{R} \rightarrow \bar{M}$  is in  $\mathcal{G}(L)$  if*

$$A_L(\bar{\gamma}|_{[-T, T]}) = \inf_{T' \in \mathbb{R}_+} h_L^{T'}(\bar{\gamma}(-T), \bar{\gamma}(T)).$$

The set  $\mathcal{G}(L)$  is nonempty. Denote by  $\bar{\gamma}_L(\cdot, \bar{m}_0, \bar{m}_1): [-T, T] \rightarrow M$  the minimizer such that  $\bar{\gamma}_L(-T) = \bar{m}_0$ ,  $\bar{\gamma}_L(T) = \bar{m}_1$  and

$$A(\bar{\gamma}_L) = \int_{-T}^T L(\bar{\gamma}_L(t), \dot{\bar{\gamma}}_L(t)) dt = \inf_{T' \in \mathbb{R}_+} h_L^{T'}(\bar{m}, \bar{m}').$$

Because of Lemma 7.1, this infimum is attained for finite  $T > 0$  if  $\bar{m}_0$  and  $\bar{m}_1$  are two different points in  $\bar{M}$ . The super-linear growth of  $L$  in  $\dot{x}$  guarantees that  $T \rightarrow \infty$  as  $-\bar{m}_{01}, \bar{m}_{11} \rightarrow \infty$ , where  $\bar{m}_{i1}$  denotes the first entry of  $\bar{m}_i$  for  $i = 0, 1$ . Given an interval  $[-T, T]$ , for sufficiently large  $-\bar{m}_{01}, \bar{m}_{11}$ , the set  $\{\bar{\gamma}_L(\cdot, \bar{m}_0, \bar{m}_1)|_{[-T, T]}\}$  is pre-compact in  $C^1([-T, T], \bar{M})$ . Let  $T \rightarrow \infty$ . By diagonal extraction argument, there is a subsequence of  $\{\bar{\gamma}_L(\cdot, \bar{m}_0, \bar{m}_1)\}$  which converges  $C^1$ -uniformly on any compact set to a  $C^1$ -curve  $\bar{\gamma}: \mathbb{R} \rightarrow \bar{M}$ . Obviously,  $\bar{\gamma} \in \mathcal{G}(L)$ , and it is shown in [C12] that

**Proposition 7.1.** *Some number  $K > 0$  exists so that  $|h_L^T(\bar{\gamma}(-T), \bar{\gamma}(T))| \leq K$  holds simultaneously for all curve  $\bar{\gamma} \in \mathcal{G}(L)$  and all  $T > 0$ .*

Each  $k \in \mathbb{Z}$  defines a Deck transformation  $\mathbf{k} : \bar{M} \rightarrow \bar{M} : \mathbf{k}x = (x_1 + k, x_2, \dots, x_n)$ . Let  $\bar{M}_\delta^- = \{x \in \bar{M} : x_1 < -\delta\}$ ,  $\bar{M}_\delta^+ = \{x \in \bar{M} : x_1 > \delta\}$ . With this notation we are able to define the set of *pseudo connecting curve* which is proved nonempty and upper semi-continuous with respect to the Lagrangian (see the proof of Lemma 6.3 and Theorem 6.3 of [C12])

**Definition 7.2** (pseudo connecting curve). *A curve  $\bar{\gamma} \in \mathcal{G}(L)$  is called pseudo connecting curve if the following holds*

$$A_L(\bar{\gamma}|_{[-T,T]}) = \inf_{\substack{T' \in \mathbb{R}_+ \\ \mathbf{k}^- \bar{\gamma}(-T) \in \bar{M}_\delta^- \\ \mathbf{k}^+ \bar{\gamma}(T) \in \bar{M}_\delta^+}} h_L^{T'}(\mathbf{k}^- \bar{\gamma}(-T), \mathbf{k}^+ \bar{\gamma}(T))$$

for each  $\bar{\gamma}(T) \in \bar{M}_\delta^-$  and  $\bar{\gamma}(T) \in \bar{M}_\delta^+$ . Denote by  $\mathcal{C}(L)$  the set of all pseudo connecting curves.

Obviously, if the space-step Lagrangian  $L$  is periodic in  $x_1$ , then a curve  $\bar{\gamma} \in \mathcal{C}(L)$  if and only if its projection  $\gamma = \pi\bar{\gamma} : \mathbb{R} \rightarrow M$  is semi-static.

*Proof of Theorem 7.2.* In autonomous system,  $\tilde{\mathcal{A}}(c)$  can be connected to  $\tilde{\mathcal{A}}(c')$  only if  $\alpha(c) = \alpha(c')$ . If  $c, c' \in \alpha^{-1}(\min \alpha)$ , then  $\tilde{\mathcal{A}}(c) \cap \tilde{\mathcal{A}}(c') \neq \emptyset$  (see [Ms]), it is trivial to connect an Aubry set to itself. So, we only need to work on the energy level set  $H^{-1}(E)$  with  $E > \min \alpha$ . Under this condition, there exists coordinate system so that  $\omega_1(\mu) > 0$  holds for each ergodic minimal measure of  $c$  and  $c'$  if they are close to each other.

The section  $\Sigma_0$  separates  $\bar{M}$  into two parts, the upper part  $\bar{M}^+$  extending to  $\{x_1 = \infty\}$  and the lower part  $\bar{M}^-$  connected to  $\{x_1 = -\infty\}$ . Denote a  $\delta$ -neighborhood of  $\Sigma_0$  in  $\bar{M}$  by  $\Sigma_0 + \delta$ , we introduce a smooth function  $\chi \in C^r(\bar{M}, [0, 1])$  such that  $\chi = 0$  if  $x \in \bar{M}^- \setminus (\Sigma_0 + \delta)$ ,  $\chi = 1$  if  $x \in \bar{M}^+ \setminus (\Sigma_0 + \delta)$ .

For those  $c'$  such that  $\langle c' - c, g \rangle = 0$  holds for each  $g \in H_1(\mathbb{T}^{j+k+\ell}, \mathbb{T}^{j+k}, \mathbb{Z})$ , there exists smooth function  $u : \bar{M} \rightarrow \mathbb{R}$  so that  $\partial u = c' - c$  if  $x \in \{|x_1| < \epsilon\} \times (O_m + \epsilon) \times \mathbb{T}^\ell$  and  $\partial u = 0$  if  $x \notin \{|x_1| < 2\epsilon\} \times (O_m + 2\epsilon) \times \mathbb{T}^\ell$ .

Without lose of generality we assume  $\tilde{\mathcal{A}}(c') \cap N_{i'} \neq \emptyset$ . We consider the set of semi-static curves which generate orbits connecting the Aubry class  $\tilde{\mathcal{A}}_{c,i}$  to another class  $\tilde{\mathcal{A}}_{c,i'}$ . The lift of the curves to  $\bar{M}$  intersect the section  $\Sigma_0$  in the set  $\text{Argmin}\{B_{c,i,i'}, \Sigma_0\}$ . We pick up a connected component of this set contained in certain tubular domain  $S_{ii'} = \{x_1 = 0\} \times O_m \times \mathbb{T}^\ell$ . Let  $\bar{\gamma}_{ii'}(t, x)$  denote the lift of semi-static curves  $\gamma_{ii'}(t, x)$  so that  $\bar{\gamma}_{ii'}(0, x) = x \in S_{ii'}$ . As all curve in  $\{\gamma_{ii'}(t, x)\}$  take  $\mathcal{A}_{c,i}$  as their  $\alpha$ -limit set and take  $\mathcal{A}_{c,i'}$  as their  $\omega$ -limit set,  $\dot{\bar{\gamma}}_{ii'}(t, x)$  is Lipschitz in  $x \in \text{Argmin}\{B_{c,i,i'}, S_{ii'}\}$ . We extend these curves to the whole  $S_{ii'}$ , so that  $\dot{\bar{\gamma}}_{ii'}(t, x)$  is still Lipschitz in  $x$  although the extended curves  $\{\bar{\gamma}_{ii'}(t, x)\}$  do not generate orbits of  $\phi_L^t$  if  $x \notin \text{Argmin}\{B_{c,i,i'}, S_{ii'}\}$ . By deforming  $\Sigma_0 \rightarrow \Sigma'$  we can assume that these curves pass transversally through the section  $\Sigma'$ .

Let  $s = s(\bar{\gamma}_{ii'}(t, x))$  denote the arc-length of the curve from  $\bar{\gamma}_{ii'}(0, x)$  to  $\bar{\gamma}_{ii'}(t, x)$  in the Euclidean metric such that  $s(\bar{\gamma}_{ii'}(0, x)) = 0$  and  $s(\bar{\gamma}_{ii'}(t, x)) > 0$  if  $t > 0$ . We approximate the function  $s$  by a smooth function  $s'$  in the tubular domain made up by the curves  $\{\bar{\gamma}_{ii'}(t, x)\}$  with  $\bar{\gamma}_{ii'}(0, x) \in S_{ii'}$ . Let  $\tau : \mathbb{R} \rightarrow [0, 1]$  be a smooth function so that  $\tau = 0$  if  $s \leq 0$ ,  $\tau = 1$  if  $s \geq s_0$  and  $\dot{\tau} > 0$  if  $s \in (0, s_0)$ . Let  $w \in C^r(T\bar{M}, [0, 1])$  such that  $w = 1$  when  $(x, \dot{x})$  is restricted in  $\{(\bar{\gamma}_{ii'}(t, x), \dot{\bar{\gamma}}_{ii'}(t, x)) : x \in S_{ii'}, s \in [0, s_0]\} + \delta$  and  $w = 0$  when  $(x, \dot{x})$  does not lie in the set  $\{(\bar{\gamma}_{ii'}(t, x), \dot{\bar{\gamma}}_{ii'}(t, x)) : x \in S_{ii'}, s \in [0, s_0]\} + 2\delta$ .

Next, we are going to show the curves in  $\mathcal{C}(L_{c,v,u})$  produce orbits of  $\phi_L^t$  connecting  $\tilde{\mathcal{A}}(c)$  to  $\tilde{\mathcal{A}}(c')$ , where the modified Lagrangian  $L_{c,v,u}$  is defined as follows

$$(7.2) \quad L_{c,v,u} = L - \langle c, \dot{x} \rangle - w \langle \partial(\tau \circ s'), \dot{x} \rangle - \chi \langle c' - c - \partial u, \dot{x} \rangle + \alpha(c),$$

where  $\chi = 0$  if  $x \in \bar{M}^- \setminus (\Sigma_0 + \delta)$ ,  $\chi = 1$  if  $x \in \bar{M}^+ \setminus (\Sigma_0 + \delta)$ , the function  $s'$  is extended to the whole space in any way one likes because  $\pi_x \text{supp } w$  is contained in the tubular domain where  $s'$  is well-defined,  $\pi_x: TM \rightarrow M$  denotes the standard projection along tangent fibers. As the first step, let us set  $\chi \equiv 0$ . Because of the upper semi-continuity of  $L \rightarrow \mathcal{C}(L)$ , each curve in  $\mathcal{C}(L_{c,v,u})$  either is contained in  $\{(\bar{\gamma}_{ii'}(t, x), \dot{\gamma}_{ii'}(t, x)) : x \in S_{ii'}\} + \delta$  or keeps away from the larger tubular domain  $\{(\bar{\gamma}_{ii'}(t, x), \dot{\gamma}_{ii'}(t, x)) : x \in S_{ii'}\} + 2\delta$ . As  $w \equiv 1$  holds on the smaller tubular domain, the term  $\langle \partial(\tau \circ s'), \dot{x} \rangle$  does not contribute to the Euler-Lagrange equation. By definition, along each orbit  $(\gamma_{ii'}(t, x), \dot{\gamma}_{ii'}(t, x))|_{s \in [0, s_0]}$  one has  $\langle \partial(\tau \circ s'), \dot{x} \rangle > 0$ . Therefore, in the lift of  $\{\gamma_{ii'}(t, x) : x \in S_{ii'}\}$ , only those curves are the member of  $\mathcal{C}(L_{c,v,u})$  for  $\chi = 0$  if they pass through the section  $S_{ii'}$ . Other curves in the lift are not in  $\mathcal{C}(L_{c,v,u})$  because they have larger Lagrange action.

Next, we recover the term  $\chi \langle c' - c - \partial u, \dot{x} \rangle$  which is  $C^2$ -small. Due to the upper semi-continuity again, all curves in  $\mathcal{C}(L_{c,v,u})$  must pass through  $S_{ii'}$  if they connect  $\mathcal{A}_{c,i}$  to  $\mathcal{A}_{c',i'}$ . As  $\partial u = c' - c$  if  $x \in \{|x_1| < \epsilon\} \times (O_m + \epsilon) \times \mathbb{T}^\ell$  and  $\partial u = 0$  if  $x \notin \{|x_1| < 2\epsilon\} \times (O_m + 2\epsilon) \times \mathbb{T}^\ell$ , one can see from the definition of  $\chi$  that the term  $\chi \langle c' - c - \partial u, \dot{x} \rangle$  does not contribute to the Euler-Lagrange equation. It implies these curves produce orbits of  $\phi_L^t$  which connects  $\tilde{\mathcal{A}}(c)$  to  $\tilde{\mathcal{A}}(c')$ .  $\square$

The orbit  $(\gamma, \dot{\gamma})$  obtained here is local minimal in following sense (analogous to 7.1):

**Local minimum:** *there exist two  $(n-1)$ dimensional disks  $V_i^-, V_{i'}^+ \subset \bar{M}$  and positive numbers  $T, d > 0$  such that  $\bar{\pi}V_i^- \subset N_i \setminus \mathcal{A}(c)$ ,  $\bar{\pi}V_{i'}^+ \subset N_{i'} \setminus \mathcal{A}(c')$ ,  $\gamma$  transversally passes  $\bar{\pi}V_i^-$  and  $\bar{\pi}V_{i'}^+$  at the time  $-T$  and  $T$  respectively, and*

$$(7.3) \quad \begin{aligned} & h_c^\infty(x^-, \bar{\pi}\bar{m}_0) + h_{L_{c,v,u}}^{T'}(\bar{m}_0, \bar{m}_1) + h_{c'}^\infty(\bar{\pi}\bar{m}_1, x^+) \\ & - \lim_{\substack{t_i^- \rightarrow \infty \\ t_i^+ \rightarrow \infty}} \int_{-t_i^-}^{t_i^+} L_{c,v,u}(\gamma(t), \dot{\gamma}(t)) dt - (t_i^- + t_i^+) \alpha(c) > 0 \end{aligned}$$

holds  $\forall (\bar{m}_0, \bar{m}_1, T') \in \partial(V_i^- \times V_{i'}^+ \times [T-d, T+d])$ ,  $x^- \in N_i \cap \pi_x(\alpha(d\gamma))$  and  $x^+ \in N_{i'} \cap \pi_x(\omega(d\gamma))$ . Where  $t_i^- \rightarrow \infty$  and  $t_i^+ \rightarrow \infty$  are the sequences such that  $\gamma(-t_i^-) \rightarrow x^-$  and  $\gamma(t_i^+) \rightarrow x^+$ .

**7.2. Local connecting orbits of type-c.** For autonomous system, if  $c'$  is equivalent to  $c$  with  $|c - c'| \ll 1$ , then  $\langle c' - c, g \rangle = 0$  holds for all  $g \in H_1(\mathcal{N}(c) \cap \Sigma_c, \mathbb{Z})$  where  $\Sigma_c$  is a section of  $M$ . So there is a function  $u$  defined on the whole torus and  $\partial u = c' - c$  holds in a small neighborhood of  $\mathcal{N}(c) \cap \Sigma_c$ . To connect  $\tilde{\mathcal{A}}(c)$  to  $\tilde{\mathcal{A}}(c')$ , we work in a coordinate system  $G_c^{-1}x$  so that  $\omega_1(\mu_c) > 0$  holds for each ergodic  $c$ -minimal measure. The new coordinate system  $G_c$  is chosen so that the lift  $\Sigma_c$  to the covering manifold  $\bar{M}$  contains infinitely many compact connected components. We fix one component, denoted by  $\Sigma_c^0$ . Other components in the lift of  $\Sigma_c$  are obtained by translating this one by  $2k\pi$  in the direction of  $x_1$ . The section  $\Sigma_c^0$  separates  $\bar{M}$  into two parts  $\bar{M}^-$  and  $\bar{M}^+$ . In  $\bar{M}^\pm$ , the coordinate  $x_1$  can be extended to  $\pm\infty$ . Let  $\text{sign}$  be a sign function defined as  $\text{sign}(x) = \pm 1$  if  $x \in \bar{M}^\pm$ .

Let  $L_{c,u}$  be a space-step Lagrangian defined on the covering manifold  $\bar{M}$

$$(7.4) \quad L_{c,u} = L - \langle c, \dot{x} \rangle - \chi \langle c' - c - \partial u, \dot{x} \rangle + \alpha(c)$$

where  $\alpha(c) = \alpha(c')$ ,  $\chi = 0$  if  $x \in \bar{M}^- \setminus (\Sigma_0 + \delta)$ ,  $\chi = 1$  if  $x \in \bar{M}^+ \setminus (\Sigma_0 + \delta)$ . Obviously, for  $c' = c$ , we have  $\bar{\pi}\mathcal{C}(L_{c,u}) = \mathcal{N}(c)$ . According to the upper semi-continuity, for sufficiently small  $|c' - c|$ , the image of each curve  $\bar{\gamma} \in \mathcal{C}(L_{c,u})$  falls in a small neighborhood of  $\mathcal{N}(c)$ . Therefore,  $\langle c' - c - \partial u, \dot{x} \rangle = 0$  holds along this curve when it passes through a small neighborhood of  $\Sigma_c^0$ . It implies that the term  $\chi \langle c' - c - \partial u, \dot{x} \rangle$  does not contribute to the Euler-Lagrange equation determined by  $\bar{L}$ . Therefore, this curve also solves the Euler-Lagrange equation for  $L$ . Clearly,  $\bar{\pi}(\bar{\gamma}(t), \dot{\bar{\gamma}}(t))$  approaches the Aubry set for class  $c'$  as  $t \rightarrow \infty$ . Therefore, we have

**Theorem 7.3** (Theorem 6.4 of [C12]: connecting orbits of type-c). *Assume the cohomology class  $c^*$  is  $c$ -equivalent to the class  $c'$  through the path  $\Gamma: [0, 1] \rightarrow H^1(\mathbb{T}^n, \mathbb{R})$ . For each  $s \in [0, 1]$ , the following are assumed:*

- (1) *there exists a coordinate systems  $G_s^{-1}x$  where the first component of rotation vector is positive,  $\omega_1(\mu_{\Gamma(s)}) > 0$  for each ergodic  $\Gamma(s)$ -minimal measure  $\mu_{\Gamma(s)}$ ;*
- (2) *for the covering space  $\bar{M}_s = \mathbb{R} \times \mathbb{T}^{n-1}$  in the coordinate system the lift of non-degenerately embedded codimension-one torus  $\Sigma_{\Gamma(s)}$  has infinitely many connected and compact components, each of which is also a codimension-one torus.*

*Then there exist some classes  $c^* = c_0, c_1, \dots, c_k = c'$  on this path, closed 1-forms  $\eta_i$  and  $\bar{\mu}_i$  on  $M$  with  $[\eta_i] = c_i$  and  $[\bar{\mu}_i] = c_{i+1} - c_i$ , and smooth functions  $\varrho_i$  on  $\bar{M}$  for  $i = 0, 1, \dots, k-1$ , such that the pseudo connecting curve set  $\mathcal{C}(L_i)$  for the space-step Lagrangian*

$$L_{c_i, u_i} = L - \langle c_i, \dot{x} \rangle - \chi_i \langle c_{i+1} - c_i - \partial u_i, \dot{x} \rangle + \alpha(c_i)$$

*possesses the properties:*

- (i) *each curve  $\bar{\gamma} \in \mathcal{C}(L_i)$  determines an orbit  $(\gamma, \dot{\gamma})$  of  $\phi_L^t$ ;*
- (ii) *the orbit  $(\gamma, \dot{\gamma})$  connects  $\tilde{\mathcal{A}}(c_i)$  to  $\tilde{\mathcal{A}}(c_{i+1})$ , i.e., the  $\alpha$ -limit set  $\alpha(d\gamma) \subseteq \tilde{\mathcal{A}}(c_i)$  and  $\omega$ -limit set  $\omega(d\gamma) \subseteq \tilde{\mathcal{A}}(c_{i+1})$ .*

**Corollary 7.1** (Corollary 6.1 of [C12]). *Let  $c_i, c_{i+1}, \chi_i$  and  $u_i$  be defined as in Theorem 7.3. Let  $U_i$  be a neighborhood of  $\mathcal{N}(c_i) \cap \Sigma_{c_i}^0$  such that  $c_{i+1} - c_i - \partial u_i|_{U_i} = 0$ . Then, there exist large  $K_i > 0, T_i > 0$  and small  $\delta > 0$  such that for each  $\bar{m}, \bar{m}' \in \bar{M}$ , with  $-K_i \leq \bar{m}_1 \leq -K_i + 2\pi, K_i - 2\pi \leq \bar{m}'_1 \leq K_i$ , the quantity  $h_{L_{c_i, u_i}}^T(\bar{m}, \bar{m}')$  reaches its minimum at some  $T < T_i$  and the corresponding minimizer  $\bar{\gamma}_i(t, \bar{m}, \bar{m}')$  satisfies the condition*

$$(7.5) \quad \text{Image}(\bar{\gamma}_i) \cap (\Sigma_{c_i}^0 + \delta) \subset U_i.$$

There is some flexibility to choose the coordinate system and the non-degenerately embedded codimension one torus. Let  $\pi_s: \bar{M}_s \rightarrow M = \mathbb{T}^n$  be a covering space such that  $\bar{M}_s = \mathbb{R} \times \mathbb{T}^{n-1}$  in the coordinate system  $G_s^{-1}x$ .

**Definition 7.3** (admissible toral section). *For  $s \in [0, 1]$ , the non-degenerately embedded codimension one torus  $\Sigma_s$  is called admissible for the coordinate system  $G_s^{-1}x$  if the lift of  $\Sigma_s$  to the covering space  $\bar{M}_s$  consists of infinitely many connected and compact components, the first component of the rotation vector is positive  $\omega_1(\mu_{\Gamma(s)})$  for each ergodic  $\Gamma(s)$ -minimal measure.*

**7.3. Existence of generalized transition chain.** Roughly speaking, a generalized transition chain is such a path  $\Gamma: [0, 1] \rightarrow H^1(M, \mathbb{R})$  that, for any  $s, s' \in [0, 1]$  with  $|s - s'| \ll 1$ , the Aubry sets  $\tilde{\mathcal{A}}(\Gamma(s))$  and  $\tilde{\mathcal{A}}(\Gamma(s'))$  are connected by local connecting orbit either of type- $h$  or of type- $c$ . An orbit  $(\gamma, \dot{\gamma})$  of the Euler-Lagrange flow  $\phi_L^t$  is said connecting two Aubry sets if the  $\alpha$ -limit set of the orbit is contained in one Aubry set and the  $\omega$ -limit set is contained in another one. It is called local connecting orbit if the first cohomology classes are close to each other.

According to a result of Bernard [Be1], both Aubry set and Mañé set are symplectic invariant. Therefore, from what have been done in previous sections one obtains clear understanding of the Aubry and Mañé sets along the resonant path. It is good enough for us to construct generalized transition chain.

Let us formulate the definition for autonomous Hamiltonian  $H: T^*M \rightarrow \mathbb{R}$  where  $M = \mathbb{T}^n$  with  $n \geq 3$ .

**Definition 7.4** (Generalized transition chain: the autonomous case). *Two cohomolgy classes  $c, c' \in H^1(M, \mathbb{R})$  are said to be joined by a generalized transition chain if there exists a path  $\Gamma: [0, 1] \rightarrow H^1(M, \mathbb{R})$  such that  $\Gamma(0) = c$ ,  $\Gamma(1) = c'$ ,  $\alpha(\Gamma(s)) \equiv E > \min \alpha$  and for each  $s \in [0, 1]$  at least one of the following cases takes place:*

- (H1) *In some finite covering manifold:  $\tilde{\pi}: \tilde{M} \rightarrow M$  the Aubry set  $\mathcal{A}(\Gamma(s))$  consists of two classes. There are two open domains  $N_i$  and  $N_{i'}$  with  $\tilde{N}_i \cap \tilde{N}_{i'} = \emptyset$ , a decomposition  $\tilde{M} = M_1 \times \mathbb{T}^\ell$ ,  $(n - \ell - 1)$ -dimensional disks  $\{O_m \subset \pi_{-1}M_1\}$  with  $\tilde{O}_m \cap \tilde{O}_{m'} = \emptyset$ , an  $(n - 1)$ -dimensional disk  $D_s$  and two small numbers  $\delta_s, \delta'_s > 0$  such that*
  - (i) *the Aubry set  $\mathcal{A}(\Gamma(s)) \cap N_1 \neq \emptyset$ ,  $\mathcal{A}(\Gamma(s)) \cap N_2 \neq \emptyset$  and  $\mathcal{A}(\Gamma(s')) \cap (N_1 \cup N_2) \neq \emptyset$  for each  $|s' - s| < \delta_s$ ,*
  - (ii)  *$\tilde{\pi}\mathcal{N}(\Gamma(s), \tilde{M})|_{D_s} \setminus (\mathcal{A}(\Gamma(s)) + \delta'_s)$  is non-empty, of which each connected component is contained in  $O_m \times \mathbb{T}^\ell$ ,*
  - (iii)  *$\langle \Gamma(s') - \Gamma(s), g \rangle = 0$  holds for each  $g \in H_1(\tilde{M}, M_1, \mathbb{Z})$ ;*
- (H2) *For each  $s' \in (s - \delta_s, s + \delta_s)$ ,  $\Gamma(s')$  is equivalent to  $\Gamma(s)$ . Some section  $\Sigma_s$  and some small neighborhood  $U$  of  $\mathcal{N}(\Gamma(s)) \cap \Sigma_s$  exist such that  $\langle \Gamma(s') - \Gamma(s), g \rangle = 0$  holds for each  $g \in H_1(U, \mathbb{Z})$ . Each class  $\Gamma(s')$  is associated with an admissible section  $\Sigma_{s'}$  for the coordinate system  $G_{s'}^{-1}x$ .*

Let us explain a bit more about the definition. In the case (H1), if the Aubry set contains only one Aubry class, one can take some finite covering  $\tilde{\pi}: \tilde{M} \rightarrow M$  so that  $\mathcal{A}(c, \tilde{M})$  contains two classes. A typical case is that  $\mathcal{A}(\Gamma(s))$  is contained in a small neighborhood of lower dimensional torus. If  $\mathcal{A}(\Gamma(s))$  contains more than one class, we choose  $\tilde{M} = M$ . The  $(n - 1)$ -dimensional disk  $D_s$  is chosen to intersect transversally all  $\Gamma(s)$ -semi static curves. If  $\ell = 0$ , there is no restriction like (iii) on the class  $\Gamma(s')$ , it turns out to be trivial.

Once such transition chain exists, one can construct diffusion orbits by variational method, which is almost the same as in [C12], recapitulated in Section 7.4. To see the existence of such transition chain in the system (1.1), let us recall that, after  $(n - 3)$  steps of approximation of frequency, a path  $\Gamma^\omega: [0, 1] \rightarrow H_1(\mathbb{T}^n, \mathbb{R})$  is fixed to pass through small neighborhood of  $\omega^1, \omega^2, \dots, \omega^k$  in any prescribed order, which admits finite covering of open segments of curves

$$\Gamma^\omega \subseteq \bigcup (\Gamma_{w,i}^\omega \cup \Gamma_{ll,i}^\omega \cup \Gamma_{s,i}^\omega \cup \Gamma_{lr,i}^\omega)$$

where  $\Gamma_{ll,i}^\omega$  connects  $\Gamma_{w,i}^\omega$  to  $\Gamma_{s,i}^\omega$ ,  $\Gamma_{lr,i}^\omega$  connects  $\Gamma_{s,i}^\omega$  to  $\Gamma_{w+1,i}^\omega$  and



- (1)  $\Gamma_{w,i}^\omega$  is obtained by reduction for single resonance. For each frequency  $\omega \in \Gamma_{w,i}^\omega$  but finitely many  $\{\omega_w^i\}$ , the Mather set  $\tilde{\mathcal{M}}(c)$  consists of only one hyperbolic periodic orbit  $(\gamma_\omega, \dot{\gamma}_\omega)$ , where  $c \in \mathcal{L}_{\beta_L}(\omega)$ ; for those  $\omega_w^i$  the Mather set consist of exactly two hyperbolic periodic orbits. All of these orbits share the same first homology class, i.e. certain  $g_i \in H_1(\mathbb{T}^n, \mathbb{Z})$  exists such that  $[\gamma_\omega] = g_i$  holds for each  $\omega \in \Gamma_{w,i}$ . These periodic orbits make up finitely many pieces of normally hyperbolic cylinder. Let  $\mathbb{C}_i = \mathcal{L}_{\beta_L}(\Gamma_{w,i}^\omega)$ , it is an open and connected channel in  $H^1(\mathbb{T}^n, \mathbb{R})$ ;
- (2) The segments  $\Gamma_{s,i}^\omega$ ,  $\Gamma_{ll,i}^\omega$  and  $\Gamma_{lr,i}^\omega$  are pretty short, lie in a small neighborhood of  $m$ -strong resonant frequency  $\omega_{s,i}$ . Their Fenchel-Legendre dual  $\mathcal{L}_{\beta_L}(\Gamma_{s,i}^\omega)$ ,  $\mathcal{L}_{\beta_L}(\Gamma_{ll,i}^\omega)$  and  $\mathcal{L}_{\beta_L}(\Gamma_{lr,i}^\omega)$  are contained in the annulus  $\mathbb{A}_i$  surrounding the flat  $\mathbb{F}_i = \mathcal{L}_{\beta_L}(\omega_{s,i})$ . The annulus  $\mathbb{A}_i$  admits a foliation of circles of cohomology equivalence, one is chosen as the chain of  $c$ -equivalence, in  $\mathcal{L}_{\beta_L}(\Gamma_{ll,i}^\omega)$  ( $l = l, r$ ) one curve is chosen as the ladder. For  $c \in \mathcal{L}_{\beta_L}(\cup \Gamma_{ll,i}^\omega \cup \Gamma_{s,i}^\omega \cup \Gamma_{lr,i}^\omega)$ , the structure of Mather set  $\mathcal{A}(c)$  may not be regular, could be complicated.

Therefore, we are able to construct a generalized transition chain  $\Gamma: [0, 1] \rightarrow H^1(\mathbb{T}^n, \mathbb{R})$

$$\Gamma \subseteq \bigcup (\Gamma_{w,i} \cup \Gamma_{ll,i} \cup \Gamma_{s,i} \cup \Gamma_{lr,i})$$

passing through small neighborhood of  $\mathcal{L}_{\beta_L}(\omega^1), \mathcal{L}_{\beta_L}(\omega^2), \dots, \mathcal{L}_{\beta_L}(\omega^k)$ , where  $\Gamma_{m,i} \subset \mathcal{L}_{\beta_L}(\Gamma_{m,i}^\omega)$  ( $m = w, s, l$ ), and

- (1)  $\Gamma_{w,i}$  passing through the interior of  $\mathbb{C}_i$ , connected to  $\Gamma_{s,i}$  by the ladder  $\Gamma_{ll,i}$  and connected to  $\Gamma_{s-1,i}$  by the ladder  $\Gamma_{lr,i-1}$ ;
- (2)  $\Gamma_{s,i}$  turns around  $\mathbb{F}_i$ , connected to  $\mathbb{C}_i$  and to  $\mathbb{C}_{i+1}$ , by the ladder  $\Gamma_{ll,i}$  and  $\Gamma_{lr,i}$  respectively.

Once such a generalized transition chain is constructed, we obtain immediately the main result of this paper (Theorem 1.1) by applying Theorem 7.4 in the following. To understand why the constructed orbit passes through the small ball  $B_\delta(p_i)$  let us note following facts

- (1) if the system is integrable, the relation  $p = c$  holds along each  $c$ -minimal orbit, there is an one to one correspondence between action variable  $p$  and frequency  $\omega \rightarrow p = \partial h^{-1}(\omega)$ . For an integrable system under small perturbation of order  $O(\epsilon)$ , it holds for each frequency  $\omega \in \mathbb{R}^n$  that the size of  $\mathcal{L}_{\beta_L}(\omega)$  is at most of order  $O(\sqrt{\epsilon})$ . Along any orbit in the Mañé set  $\tilde{\mathcal{M}}(c)$ , the action variable is restricted in the  $O(\sqrt{\epsilon})$ -ball centered at  $c$  if we treat  $c$  and  $p$  as points in  $\mathbb{R}^n$ .
- (2) keep it in mind that  $\omega^i$  is chosen close to  $\partial h(p^i)$  and the transition chain passes through  $\mathcal{L}_{\beta_L}(\omega^i)$ . From the construction of diffusion orbits, one can see that there is a quite long period when the orbit keeps very close to certain orbit in the Mañé set  $\tilde{\mathcal{M}}(c)$  with  $c \in \mathcal{L}_{\beta_L}(\omega^i)$ . It implies the orbit passing through neighborhood of  $p^i$ .

**7.4. Variational construction of globally connecting orbits.** It was done in [C12] for general dimension  $n$ . Here we only recapitulate the main steps in the proof and refer readers to [C12] for details.

**Theorem 7.4** (Existence of globally connecting orbit). *If  $c$  is connected to  $c'$  by a generalized transition chain  $\Gamma$  as in Definition 7.4, then*

- (1) *there exists an orbit of the Lagrange flow  $\phi_L^t$ ,  $(\gamma, \dot{\gamma}): \mathbb{R} \rightarrow TM$  which connects the Aubry set  $\tilde{\mathcal{A}}(c)$  to  $\tilde{\mathcal{A}}(c')$ , namely,  $\alpha(d\gamma) \subseteq \tilde{\mathcal{A}}(c)$  and  $\omega(d\gamma) \subseteq \tilde{\mathcal{A}}(c')$ ;*
- (2) *for any  $c_1, c_2, \dots, c_k \in \Gamma$  and arbitrarily small  $\delta > 0$ , there exists time-sequence  $t_1 < t_2 < \dots < t_k$  such that the orbit  $(\gamma, \dot{\gamma})$  of  $\phi_L^t$  passes through  $\delta$ -neighborhood of the Aubry set  $\tilde{\mathcal{A}}(c_i)$  at the time  $t = t_i$ .*

*Sketch of the proof.* The proof of this theorem is the same as the proof of Theorem 7.1 of [C12]. We only sketch the idea of the proof here, readers can refer to [C12] for the details. Because of the condition of generalized transition chain, there is a sequence  $0 = s_0 < s_1 < \dots < s_k = 1$  such that for each  $0 \leq j < k$ ,  $\tilde{\mathcal{A}}(\Gamma(s_j))$  is connected to  $\tilde{\mathcal{A}}(\Gamma(s_{j+1}))$  by local minimal orbit either of type- $h$  with incomplete intersection or of type- $c$ . The global connecting orbits are constructed shadowing such a sequence of orbits.

Recall the construction of local connecting orbit as above, for each  $i \in \{0, 1, \dots, k\}$  let  $\eta_i(x, \dot{x}) = \langle c_i, \dot{x} \rangle$  and

$$\mu_i(x, \dot{x}) = w_i \langle \partial(\tau_i \circ s_i'), \dot{x} \rangle, \quad \psi_i(x, \dot{x}) = \chi_i \langle c_{i+1} - c_i - \partial u_i, \dot{x} \rangle$$

in certain coordinate system  $G_i^{-1}x$  (see (7.2), (7.4) for the definition), if it is for type- $c$ , we set  $\mu_i = 0$ . For each integer  $k$  we introduce a translation operator on functions  $k^*f(x_1, x_2, \dots, x_n) = f(x_1 - k, x_2, \dots, x_n)$ .

Let  $\tilde{\pi}: \mathbb{R}^n \rightarrow M$  be the universal covering space. For a curve  $\tilde{\gamma}: [-K, K'] \rightarrow \mathbb{R}^n$ , let  $\gamma = \tilde{\pi}\tilde{\gamma}: [-K, K'] \rightarrow M$ . Let  $\vec{t} = (t_0^-, t_1^\pm, \dots, t_{k-1}^\pm, t_k^+)$ ,  $\vec{x} = (\tilde{x}_0^-, \tilde{x}_1^\pm, \dots, \tilde{x}_{k-1}^\pm, \tilde{x}_k^+)$  with  $t_i^+ < t_i^- < t_{i+1}^+$ ,  $t_0^- = -K$  and  $t_k^+ = K'$ . we consider the minimal action

$$\begin{aligned} h_L^{K, K'}(m, m', \vec{x}, \vec{t}) = & \inf \sum_{i=0}^k \int_{t_i^+}^{t_i^-} (L - \eta_i)(d\tilde{\gamma}_i^-(t)) dt \\ & + \sum_{i=0}^{k-1} \int_{t_i^-}^{t_{i+1}^+} (L - \eta_i - (k_i G_i)^*(\mu_i + \psi_i))(d\tilde{\gamma}_i^+(t)) dt \end{aligned}$$

where the infimum is taken over all absolutely continuous curves  $\tilde{\gamma}: [-K, K'] \rightarrow \mathbb{R}^n$  satisfying the boundary conditions  $\tilde{\gamma}^-(t_i^-) = \tilde{\gamma}_i^-(t_i^-) = \tilde{x}_i^-$ ,  $\tilde{\gamma}_i^+(t_{i+1}^+) = \tilde{x}_{i+1}^+$  for  $i = 0, 1, \dots, i_{k-1}$ ,  $\gamma(-K) = m$ ,  $\gamma(K') = m'$ . By carefully setting boundary condition we find that the minimizer is smooth everywhere, along which the term  $(k_i G_i)^*(\mu_i + \psi_i)$  does not contribute to the Euler-Lagrange equation. It is guaranteed by the local minimality of (7.5) as well as (7.3) and setting the translation  $k_{i+1} - k_i$  sufficiently large. The condition of incomplete intersection does not cause new difficulty in verifying the smoothness of the minimizer. Therefore, the minimizer produces an orbit  $(\tilde{\gamma}, \dot{\tilde{\gamma}})$  of  $\phi_L^t$  which has the properties stated in the theorem.  $\square$

## 8. GENERICITY AND THE PROOF OF THE MAIN THEOREM

In this section, we first describe the generic set of perturbations in the  $C^r$ -function space, for each  $\epsilon P$  in this set the Hamiltonian  $h(y) + \epsilon P(x, y)$  admits diffusing orbits. This set has a multiple filtration cusp-residual structure. After that we complete the proof of the main theorem.

### 8.1. Generic conditions: the multi-filtered cusp-residual condition (MFCR).

The construction of diffusion orbits requires a list of conditions on perturbations, we shall verify them one by one in this section. There are mainly two types of conditions of genericity. The first type is called generic condition on hyperbolicity (HGC) since they are used to guarantee normal hyperbolicity of NHICs. The second type is called generic conditions on transversality (TGC), since they are used to guarantee certain “transversal” intersection of the stable and unstable sets of the Aubry sets.

Essentially, the genericity problem is reduced to the following two parametric transversality (PT) results, PT1 for HGC and PT2 for TGC.

- (PT1) Consider a function  $F_\zeta(x)$ ,  $M \rightarrow \mathbb{R}$ ,  $M$  compact manifold. Suppose  $F$  is  $C^r$ ,  $r \geq 4$  in  $x$  and Lipschitz in  $\zeta \in [\zeta_0, \zeta_1]$ , then for  $V(x)$  in an open and dense subset of  $C^r(M)$ , the global max of  $F_\zeta(x) + V(x)$  is non degenerate uniformly for all  $\zeta$  except for finitely many  $\zeta_i$ 's, the function  $F_{\zeta_i}(x)$  has two nondegenerate global maxima.
- (PT2) Consider a function  $F_\zeta(x)$ ,  $M \rightarrow \mathbb{R}$ . Suppose  $F$  is Lipschitz in  $x$  and  $\alpha$  Hölder in  $\zeta \in [\zeta_0, \zeta_1]$ , then for generic  $V(x) \in C^r(M)$ ,  $r \geq 0$ , the function  $F_\zeta(x) + V(x)$  is non constant for all  $\zeta$ .

PT1 is proved in [CZ1, CZ2] and used to establish the genericity of  $(\mathbf{H1})_j$  and  $(\mathbf{H2})_j$ ,  $j = 1, 2, \dots, n-2$ . PT2 is proved in [CY1, CY2].

8.1.1. *Generic condition on hyperbolicity (HGC) and multi-filtered generic set.* We have two types of HGCs:  $(\mathbf{H1})_j$  and  $(\mathbf{H2})_j$ ,  $j = 1, 2, \dots, n-2$  (see Section 4.1 and 4.2.2). The HGC of the  $j = 1$  case is in Section 3.2 and 3.4 (see  $(\mathbf{H1})$ ,  $(\mathbf{H2.1})$ ,  $(\mathbf{H2.2})$ , Theorem 3.4 and Lemma 3.2, we do not include analogues of  $(\mathbf{H2.1})$  and  $(\mathbf{H2.2})$  in  $(\mathbf{H2})_j$  since the spectral gap is easily guaranteed in the following reduction of order and will not affect the multiple-filtration structure). The HGC of the  $j = 2$  case is in Section 3.11 (see  $(\mathbf{H1})_2$ ,  $(\mathbf{H2})_2$  there).

Since we have finitely many such HGCs, each of which is shown to be generic, we get a residual set  $\mathfrak{V}$  in  $C^r(T^*\mathbb{T}^n)$  consists of  $P(x, y)$  satisfying all the above HGCs. We point out that such a residual set also exists in  $C^r(\mathbb{T}^n)$  (Mañé genericity) because in this case  $(\mathbf{H1})_j$  is trivial and  $(\mathbf{H2})_j$  is always shown to be generic in  $C^r(\mathbb{T}^n)$  as stated. In the following, we shall only talk about  $C^r(T^*\mathbb{T}^n)$  keeping in mind that all the arguments hold also for  $C^r(\mathbb{T}^n)$ .

The generic set  $\mathfrak{V}$  has a multi-filtered structure, namely, there is a filtration  $\mathfrak{V} = \cup_{i_1 \geq 1} \mathfrak{V}_{i_1}^1$  with  $\mathfrak{V}_{i_1}^1 \subset \mathfrak{V}_{i_1+1}^1$  and each  $\mathfrak{V}_{i_1}^1$  admits a new filtration  $\mathfrak{V}_{i_1}^1 = \cup_{i_2 > i_1} \mathfrak{V}_{i_1, i_2}^2$ ,  $\mathfrak{V}_{i_1, i_2}^2 \subset \mathfrak{V}_{i_1, i_2+1}^2$  and so on. After  $n-2$  levels of filtration, we have  $\mathfrak{V} = \cup_{i_1 < \dots < i_{n-2}} \mathfrak{V}_{i_1, \dots, i_{n-2}}^{n-2}$ . Such a multiple filtration can be defined for any set if triviality is allowed, e.g.  $\mathfrak{V}_{i_1}^1 = \mathfrak{V}$  for all  $i_1$  in the first level of filtration. To avoid such triviality, we require the inclusion  $\mathfrak{V}_{i_1, \dots, i_{j-1}, i_j}^j \subset \mathfrak{V}_{i_1, \dots, i_{j-1}, i_j+1}^j$  to be proper for all  $i_j$  large enough and for all  $j = 1, 2, \dots, n-2$ . The multi-filtration has a natural lexicographical partial order, given by  $\mathfrak{V}_{i_1, \dots, i_{n-2}}^{n-2} \subset \mathfrak{V}_{i'_1, \dots, i'_{n-2}}^{n-2}$  if  $(i_1, \dots, i_{n-2}) \preceq (i'_1, \dots, i'_{n-2})$ , i.e.  $i_j \leq i'_j$  for all  $j = 1, 2, \dots, n-2$ .

We denote by  $\mathfrak{V}^1$  the set of functions in  $C^r(T^*\mathbb{T}^n)$  satisfying  $(\mathbf{H1})$ ,  $(\mathbf{H2.1})$ ,  $(\mathbf{H2.2})$ , Theorem 3.4 and Lemma 3.2. i.e. HGC for the first step of reduction of order. Notice that these conditions are only imposed on Fourier modes of  $P(x, y)$  in  $\text{span}_{\mathbb{Z}}\{\mathbf{k}'\}$  or  $\text{span}_{\mathbb{Z}}\{\mathbf{k}', \mathbf{k}''\}$  where  $\langle \mathbf{k}', \omega_a \rangle = 0$  for all  $a$  and  $\langle \mathbf{k}'', \omega_a \rangle = 0$  for finitely many

$a$ 's. We next decompose  $\mathfrak{V}^1 = \cup_{i_1=1}^{\infty} \mathfrak{U}_{i_1}^1$  with  $\mathfrak{U}_{i_1}^1 \subset \mathfrak{U}_{i_1+1}^1$ , where  $\mathfrak{U}_{i_1+1}^1$  is the set of  $P(x, y) \in C^r(T^*\mathbb{T}^n)$  satisfying

- $-\partial_{x_2'}^2 Z'(x_2^*, y) > 1/i_1^2$  in **(H1)** in Section 3.2;
- $\lambda_2 > \lambda_1 + 1/i_1$  in **(H2.1)** of Section 3.4;
- The normal Lyapunov exponents of the periodic orbits given by Theorem 3.4 and Lemma 3.2 have absolute values greater than  $1/i_1$ .

For  $P(x, y) \in \mathfrak{U}_{i_1}^1$ , we get a NHIC of dimension  $2n - 2$  whose normal hyperbolicity is quantified by  $1/i_1$ .

This is the first level of filtration. We next describe the second level of filtration. For each  $\mathfrak{U}_{i_1}^1$  there is an associated  $\Delta_{i_1}^1 \geq 0$  ( $\Delta_{i_1}^1 = 0$  iff  $\mathfrak{U}_{i_1}^1 = \emptyset$ ) such that for all  $\mathbf{k} \in \mathbb{Z}^n$  with  $|\mathbf{k}|^{-r} < \Delta_{i_1}^1$ , we have that the NHIM Theorem 3.1 and 3.2 are applicable to show the persistence of NHIC when the perturbation is  $\delta \bar{Z}_2(\langle \mathbf{k}', x \rangle, \langle \mathbf{k}'', x \rangle, \langle \mathbf{k}, x \rangle, y)$  which consists of those Fourier modes in  $P(x, y)$  in  $\text{span}_{\mathbb{Z}}\{\mathbf{k}', \mathbf{k}'', \mathbf{k}\} \setminus \text{span}_{\mathbb{Z}}\{\mathbf{k}', \mathbf{k}''\}$ . We choose one such  $\mathbf{k}$  denoted by  $\mathbf{k}^2$  which is perpendicular to  $\omega_a^2$  in (3.38) and (4.4). Here both  $\mathbf{k}^2$  and  $\omega_a^2$  depends on  $i_1$  through  $\Delta_{i_1}^1$ . For this  $\mathbf{k}^2$ , we introduce a residual set  $\mathfrak{V}_{i_1}^1 \subset \mathfrak{U}_{i_1}^1$  so that functions in  $\mathfrak{V}_{i_1}^1$  satisfy **(H1)**<sub>2</sub>, **(H2)**<sub>2</sub>. Each  $\mathfrak{V}_{i_1}^1$  admits a further filtration  $\mathfrak{V}_{j_1}^1 := \cup_{i_2 > i_1} \mathfrak{U}_{i_1, i_2}^2$  where  $\mathfrak{U}_{i_1, i_2}^2$  consists of functions in  $\mathfrak{V}_{i_1}^1$  satisfying further

- The eigenvalues of the matrix  $-\text{Hess}Z_2'$  in **(H1)**<sub>2</sub> are greater than  $1/i_2^2$ .
- The normal Lyapunov exponents of the periodic orbits given by **(H2)**<sub>2</sub> have absolute values greater than  $1/i_2$ .

Perturbation  $P \in \mathfrak{U}_{i_1, i_2}^2$  admits a NHIC of dimension  $2(n - 2)$  whose normal Lyapunov exponents quantified by  $1/i_2$ . As before, for each  $\mathfrak{U}_{i_1, i_2}^2$  there is an associated number  $\Delta_{i_1, i_2}^2 \geq 0$  ( $\Delta_{i_1, i_2}^2 = 0$  iff  $\mathfrak{U}_{i_1, i_2}^2 = \emptyset$ ) which gives the upper bound for  $|\mathbf{k}^3|^{-r}$  where  $\mathbf{k}^3$  is perpendicular to  $\omega_a^3$  in (4.4) such that Fourier modes containing  $\langle \mathbf{k}^3, x \rangle$  does not destroy the NHIC when applying the NHIM Theorem 3.1 and 3.2. Once this  $\mathbf{k}^3$  is chosen, we apply **(H1)**<sub>3</sub> and **(H2)**<sub>3</sub>.

We repeat this procedure for  $n - 2$  steps until all the HGCs **(H1)**<sub>j</sub>, **(H2)**<sub>j</sub> are applied to get a residual set  $\mathfrak{V} = \cup_{i_1 < \dots < i_{n-2}} \mathfrak{V}_{i_1, \dots, i_{n-2}}^{n-2}$  in  $C^r(T^*\mathbb{T}^n)$  that admits  $(n - 2)$  levels of filtration structure. For each  $P \in \mathfrak{V}$ , the system (1.1) admits NHIC of dimension four (for the flow and two for the Poincaré map). See Theorem 5.1.

The term “cusp residual” can be understood as follows. On the one hand, for fixed  $\varepsilon$ , we cannot consider all  $P \in \mathfrak{V}$  since there is an  $\varepsilon$  remainder (the  $R_{II}$  term) in the Hamiltonian which may destroy the NHIC if the normal hyperbolicity is too weak. Hence for each fixed  $\varepsilon$ , we consider only  $P \in \mathfrak{V}_{i_1, \dots, i_{n-2}}^{n-2}$  for some labeling of  $(i_1, \dots, i_{n-2})$ , which is not residual in  $C^r(T^*\mathbb{T}^n)$  since its complement contains open sets. We call such a set  $\mathfrak{V}_{i_1, \dots, i_{n-2}}^{n-2}$  a cusp set. On the other hand, for each  $P \in \mathfrak{V}$ , there is  $\varepsilon_P$  sufficiently small such that the NHIC persists the  $\varepsilon$  remainder for all  $\varepsilon < \varepsilon_P$ . However, this  $\varepsilon_P$  cannot be uniformly bounded away from 0 for all  $P \in \mathfrak{V}$ .

**8.1.2. The transversality genericity condition (TGC) in the presence of NHIC.** For each  $P \in \mathfrak{V}$ , the MFCR set, the HGCs are all satisfied. We can find  $\varepsilon$  small enough such that NHIC exists in a large part of the phase space applying Theorem 5.1. The NHIC containing all the Aubry sets  $\tilde{A}(c)$  with  $c \in \mathcal{L}_\beta(\omega_a^\sharp)$  in the part of phase space  $\gamma$  away from the strong double resonance, where  $\omega_a^\sharp = \lambda_a^\sharp(a, \frac{p_2}{q_2}, \frac{p_3}{q_3}, \dots, \frac{p_n}{q_n})$  is in (4.8).

In the homogenized system, the lengths of NHIC that we get are of order  $O(\varepsilon^{-1/2})$ . From now on, we consider  $\varepsilon$  as fixed.

In the presence of NHIC, the problem of constructing diffusing orbit is well understood as *a priori* unstable systems in [CY1, CY2]. We recapitulate the main ideas of the genericity proof of TGC in [CY1, CY2].

Each Aubry set  $\tilde{\mathcal{A}}(c)$ ,  $c \in \mathcal{L}_\beta(\omega_a^\sharp)$  carries stable and unstable “manifolds” whose graphs are  $(x, d_x u_c^\pm + c)$  in terms of weak KAM solutions  $u_c^\pm$ . The construction of diffusing orbit requires either (H1) or (H2) in Definition 7.4 holds. (H2) holds when  $\tilde{\mathcal{A}}(c)$  is not an invariant curve, so it remains to satisfy (H1) when for all  $\tilde{\mathcal{A}}(c)$  being curves. The intersection of the stable and unstable manifold are given by  $du_c^- - du_c^+ = 0$  and the projected Mañé set  $\mathcal{N}(c) = \text{Argmin} B_c(x)$  where the barrier function  $B_c(x) = u_c^- - u_c^+$ . So to satisfy (H2)(ii) in Definition 7.4, the problem is reduced to showing that  $B_c(x)$  is non constant uniformly for all  $c$  with  $\tilde{\mathcal{A}}(c)$  being an invariant curve. The parametric transversality result PT2 applies if we can show that the family of functions  $B_c(x)$  can be parametrized by certain parameter with respect to which the function is Hölder. It is shown in [CY1, CY2] that  $B_\sigma(x)$  parametrized by a parameter  $\sigma$  is  $1/2$ -Hölder, i.e.  $\|B_{\sigma_1} - B_{\sigma_2}\|_{C^0} \leq C|\sigma_1 - \sigma_2|^{1/2}$  for  $\sigma_1, \sigma_2$  corresponding to invariant curves. In our setting this estimate holds in the homogenized system and the constant  $C$  becomes  $C\varepsilon^{1/2}$  in the original scale. Since  $\varepsilon$  is now a constant and only the Hölder exponent is relevant in the proof of PT2, we conclude by applying the argument of [CY1, CY2] that

for each cusp set  $\mathfrak{Y}_{i_1, \dots, i_{n-2}}^{n-2} \in C^r(T^*\mathbb{T}^n)$ , there exists  $\varepsilon_0$  such that for each  $\varepsilon < \varepsilon_0$ , there exists a residual set  $\mathfrak{X}_{i_1, \dots, i_{n-2}}^{n-2} \subset \mathfrak{Y}_{i_1, \dots, i_{n-2}}^{n-2}$  such that the Hamiltonian system (1.1) with perturbation in  $P \in \mathfrak{X}_{i_1, \dots, i_{n-2}}^{n-2}$  satisfies Definition 7.4 along the NHIC given by Theorem 5.1.

It is shown in [C12, C15b] that the same statement holds if the function space  $C^r(T^*\mathbb{T}^n)$  is replaced by  $C^r(\mathbb{T}^n)$ .

**8.1.3. TGC for cohomology equivalence.** It is a condition for  $V_2(\tilde{x})$  in (4.19) if we encounter the 1<sup>st</sup>-strong resonance. For generic  $V_2$ , it holds for each  $\tilde{c} \in \partial\mathbb{F}_0$  that the Mañé set of the system  $\frac{1}{2}\langle A\tilde{y}, \tilde{y} \rangle + V_2$  does not cover the 2-torus:  $\tilde{\pi}\mathcal{N}(\tilde{c}, \hat{0}) \subsetneq \mathbb{T}^2 \ni (x_1, x_2)$ . It was shown in Theorem 6.1 (see also Theorem 5.1 of [C12]).

**8.1.4. TGC for ladder construction.** A ladder  $\mathbb{L}$  is constructed connecting a circle of cohomology equivalence to an  $s$ -resonant channel. This ladder is composed of  $m - 2$  simple ladders if we are encountered  $m$ -strong resonance

$$\mathbb{L} = \mathbb{L}_1 * \mathbb{L}_{k+1} * \dots * \mathbb{L}_{m-2}.$$

Each  $\mathbb{L}_k$  is constructed as a transition chain, i.e. any two Aubry sets  $\tilde{\mathcal{A}}(c)$ ,  $\tilde{\mathcal{A}}(c')$  are connected by local minimal orbit of type- $h$  if  $c, c' \in \mathbb{L}_k$  are close to each other. As explained in Remark 6.2, this is an autonomous version of Arnold mechanism. Hence the construction of the diffusing orbit is similar to the *a priori* unstable case and the genericity of (TGC) is also similar except that we need to use a new regularity result for weak KAM solutions.

We need to check the conditions for  $\mathbb{L}_\ell$  twice. First, as the potential in (4.19) takes the form  $\sum_{i=2}^{\ell+2} \delta_i V_i(x_1, \dots, x_i)$ . Second, when the potential eventually takes the form  $\sum_{i=2}^m \delta_i V_i(x_1, \dots, x_i)$ .

- At the beginning of constructing  $\mathbb{L}_\ell$ , we have a reduced system with  $(n - \ell)$  degrees of freedom, where the last  $n - \ell - 2$  coordinates are cyclic, namely, the Hamiltonian under consideration is independent of  $(x_{\ell+3}, \dots, x_{n-1})$ . It allows us to treat this system as two degrees of freedom so that theorem of regularity properties of weak KAMs of system with two degrees of freedom (Theorem 9.2) can be applied. As the reduced system lies on normally hyperbolic manifold, such regularity of elementary weak KAMs can be extended to the whole space (Theorem 9.3). Using techniques developed in Section 8.3 of [C12] (see Theorem 8.1 there), one can show it is a generic requirement for  $V_\ell$  that each connected component of the set of minimal points of  $B_c$

$$\text{Argmin}\{B_c, \Sigma_c \cap (\mathbb{T}^{n-\ell} \setminus \mathcal{A}(c) + \epsilon)\}$$

is contained in  $O_m \times \mathbb{T}^{n-\ell-2}$ , where  $\Sigma_c = \Sigma_{c,3} \times \mathbb{T}^{n-\ell-2}$  is codimension one section,  $\Sigma_{c,3}$  is a section in  $\mathbb{T}^3$  which is the configuration space for the first three coordinates,  $O_m \subset \Sigma_{c,3}$  is a 2-dimensional disk.

- After the construction of all simple ladders,  $\mathbb{L}_1, \dots, \mathbb{L}_{m-2}$ , we need to check whether  $\mathbb{L}_1, \dots, \mathbb{L}_{m-3}$  can survive small perturbation. For  $\mathbb{L}_\ell$ , the extra perturbation takes the form of  $\sum_{i=\ell+1}^{n-1} \delta_i V_i(x_1, \dots, x_i)$ . Due to the upper-semi continuity of Mañé set in perturbation, it holds  $\forall c \in \mathbb{L}_\ell$  that each connected component of

$$\text{Argmin}\{B_c, \Sigma_c \cap (\mathbb{T}^{n-\ell} \setminus \mathcal{A}(c) + \epsilon)\}$$

is still contained in  $O_m \times \mathbb{T}^{n-\ell-2}$  provided  $\delta_i \ll \delta_\ell$  for  $i \in \{\ell+1, \dots, m-2\}$ . As a matter of fact, certain small constants  $d > 0, D(O_m, d) > 0$  exists such that

$$\min_{\partial(O_m+d) \times \mathbb{T}^{n-\ell+3}} B_c - \min_{\Sigma_c} B_c > D(O_m, d).$$

If  $|\sum_{i=\ell+1}^m \delta_i V_i(x_1, \dots, x_i)| \leq \Delta_\ell^\dagger(D(O_m, d))$  with sufficiently small  $\Delta_\ell^\dagger > 0$ , the property as above persists.

**8.1.5. The MFCR set in the main theorem.** We get a residual set  $\mathfrak{V}$  by imposing the HGC and further a set residual in  $\mathfrak{V}$  (hence also residual in  $C^r(T^*\mathbb{T}^n)$  (or  $C^r(\mathbb{T}^n)$ )) by taking union over all the sets  $\mathfrak{R}_{i_1, \dots, i_{n-2}}^{n-2}$  such that both HGC and TGC are satisfied. Perturbations chosen in the last residual set have Definition 7.4 satisfied hence admit diffusing orbits. We then apply the Kuratowski-Ulam theorem to get a residual set  $\mathfrak{R}$  on the unit sphere  $\mathfrak{S}_1$  and for each  $P \in \mathfrak{R}$  a number  $\varepsilon_P$  and a set  $R_P$  residual in  $[0, \varepsilon_P]$  such that  $\{\lambda P : P \in \mathfrak{R}, \lambda \in R_P\}$  satisfy both HGC and TGC hence admit diffusing orbits. This gives the cusp-residual set in the statement of the main theorem.

**8.2. Proof of the main theorem.** Given any two  $\rho$ -balls as stated in the theorem in the space of action variables, we find a  $\rho$ -accessible path (c.f. Definition 4.2) according to Theorem 4.1. It is enough to show how to build the generalized transition chain (c.f. Definition 7.4) along such a  $\rho$ -accessible path. We show in Section 7.4 that such a generalized transition chain gives rise to an orbit of the Hamiltonian system visiting the Aubry sets associated to the cohomology classes along the transition train.

Since the  $\rho$ -accessible path is piecewise linear, we distinguish two different cases:

- 1, along each admissible line segment,
- 2, transition from one line to another.

Here we replace  $\Delta(\mathbf{k}'^j)$  in the item (2) in Definition 4.1 of the admissible line segment by  $\min\{\Delta(\mathbf{k}'^j), \Delta_j^\dagger\}$ , where  $\Delta_j^\dagger$  is given in the second bullet point of Section 8.1.4. With this change in the definition of admissible frequency line segment, the proof of Theorem 4.1 goes through.

For the case 1, along each admissible line segment, we have further two cases

- 1.1, Single resonance and weak double resonance,
- 1.2, Strong double resonance appearing at the  $j$ -th step of the reduction of order.

For the case 1.1, we have a complete understanding of the structure of the Aubry set in Theorem 5.1. This case is the same as systems of *a priori* unstable type studied in [CY1, CY2]. The generalized transition chain is built by applying Theorem 7.2 with  $\ell = 0$  (complete transversal intersection).

For the case 1.2, the transition chain is built step by step as we did in Section 5 and 6.2. We consider the case  $j = 0$ , i.e. we encounter the strong double resonance before any reduction of order, and the cases with  $j > 0$  are done similarly. The method is to apply the  $c$ -equivalence in Lemma 6.1 in Section 6.2 to move the first two components of the cohomology classes. Next, we perform the scheme of ladder climbing in Section 6.2. after each reduction of order. Here after  $j$ -th reduction of order, we have a system of  $n - j$  degrees of freedom. In this case, we apply Theorem 7.2 with  $\ell = n - j - 2$  being the dimensions whose transversal intersection of the stable and unstable “sets” of the Aubry sets are not needed. This completes the construction of the generalized transition chain for case 1.

Next we consider the case 2. Consider frequency vector of the form  $(a, \frac{p}{Q}, \frac{p}{q}, \hat{\omega}_{n-3}^*)$ . According to the construction in Theorem 4.1, when we stop moving  $a$  and start moving the second component denoted by  $b$ , we set  $a$  to be a rational number that is very close to a Diophantine number  $\omega_1^{*f}$ ,  $|a - \omega_1^{*f}| \leq \mu_{n-2}$ . We next perform the procedure of reduction of order for  $n-3$  steps to arrive at a frequency vector of the form  $(a, b, \frac{p_n}{q_n})$ , where  $|\frac{p_n}{q_n} - \omega_n^*| \leq \mu_{n-3}$ , associated to a Hamiltonian system of three degrees of freedom. Since all the three components of the frequency vector is rational, we are in the case of double resonance for a Hamiltonian system of three degrees of freedom. Suppose  $a/(p_n/q_n) := \frac{p_a}{q_a}$  and  $b/(p_n/q_n) := \frac{p_b}{q_b}$  with  $\text{g.c.d.}(p_a, q_a) = 1$  and  $\text{g.c.d.}(p_b, q_b) = 1$ , and we consider resonance vectors of the form  $\mathbf{k}_a := (q_a, 0, -p_a)$ ,  $\mathbf{k}_b := (0, q_b, -p_b)$ . The frequency line obtained by varying  $a$  (respectively  $b$ ) lies in the kernel of  $\mathbf{k}_b$  (respectively  $\mathbf{k}_a$ ). Applying Proposition 2.2 for double resonances and perform linear symplectic transformations, we reduce the Hamiltonian system into the form of  $\tilde{\mathbf{G}}$  in Lemma 3.1 with a remainder that can be chosen to be as small as we wish. We next apply the method of  $c$ -equivalence in Lemma 6.1 in Section 6.2 to build transition chain to transit between two frequency lines.

For finitely many  $\rho$ -balls in  $\alpha^{-1}(E)$ , a transition chain can be constructed similarly to pass through these balls. It gives rise to an orbit of the Hamiltonian system visiting the Aubry sets associated to the cohomology classes along the transition chain, namely, the constructed orbits visit these balls  $B_\rho(y_i)$  in any prescribed order. The smoothness requirement  $r > 2n$  is given in Remark 4.1. This completes the proof of the main result (Theorem 1.1).

## 9. PARAMETRIZATION OF WEAK KAM SOLUTIONS AND ITS MODULUS OF CONTINUITY

In this section, we show that the weak KAM solutions on a level set of a Tonelli system of two degrees of freedom can be parametrized by a parameter such that they have Hölder modulus of continuity. This is a substitute of the Hölder regularity result of [CY1] to show the genericity of the perturbation  $P$  needed for the construction of ladders in Section 6.2.

**9.1. Topology of level set of weak KAM solutions.** The modulus of continuity of weak KAM solution is obtained for some “volume” parameter. The introduction of this parameter relies on the well-understanding of the topology of level set of weak KAM solutions. Denote the level set of  $U - U'$  by

$$Z_{U,U'} = \{x \in \mathbb{R}^n : U(x) - U'(x) = 0\},$$

and let

$$\Omega_{U,U'} = \{x \in \mathbb{R}^n : U(x) > U'(x)\}.$$

We let  $U_c = \bar{u}_c + \langle c, x \rangle$  and drop the subscript  $c$  when it is clearly indicated, where  $\bar{u}_c$  is a globally elementary weak KAM solution for certain Aubry class in  $\mathcal{A}(c)$  (see Appendix A.3 for its definition). Obviously,  $U_c$  is a viscosity solution of the Hamilton-Jacobi equation

$$H(x, \partial_x u) = \alpha(c), \quad x \in \mathbb{R}^n.$$

Global viscosity solutions exist provided  $\alpha$  is not smaller than so-called critical value, i.e. the minimum of the  $\alpha$  function.

**Theorem 9.1.** *Let  $U = u + \langle c, x \rangle$ ,  $U' = u' + \langle c', x \rangle$  be globally elementary weak KAM solutions, where  $c$  and  $c'$  are the corresponding co-homology classes,  $u$  and  $u'$  are bounded on the whole  $\mathbb{R}^n$ , then*

*If  $\alpha(c) > \alpha(c')$  then  $\Omega_{U',U}$  is simply connected and unbounded,  $\Omega_{U,U'}$  contains only one unbounded connected component;*

*If  $\alpha(c) = \alpha(c') > \min \alpha$ ,  $c$  and  $c'$  are not in one flat, then both  $\Omega_{U,U'}$  and  $\Omega_{U',U}$  are simply connected and unbounded.*

Before the proof, let us state a lemma. A function  $m: [0, \infty) \rightarrow [0, \infty)$  is called a modulus if it is continuous, non-decreasing and satisfies  $m(0) = 0$ .

**Lemma 9.1.** *For the Hamiltonian  $H$  we assume that there is a modulus  $m$  such that*

$$H(y, \lambda(x - y)) - H(x, \lambda(x - y)) \leq m(\lambda|x - y|^2 + |x - y|)$$

*for  $\lambda \geq 0$  and  $x, y \in \Omega$ . Let  $\Omega$  be an open subset of  $\mathbb{R}^n$  and  $f \in C(\Omega)$  satisfy  $f(x) < 0$  for  $x \in \Omega$ . Let  $u, v \in C(\bar{\Omega})$  satisfy*

$$H(x, Du) \leq f(x) \quad \text{and} \quad H(x, Dv) \geq 0 \quad \text{in } \Omega$$

*in the viscosity sense. If  $\partial\Omega \neq \emptyset$ ,  $u \leq v$  on  $\partial\Omega$ ,  $u - v$  is bounded in  $\Omega$ , then  $u \leq v$  on  $\Omega$ .*

*Let  $u, v \in C(\bar{\Omega})$  satisfy*

$$H(x, Du) \leq 0 \quad \text{and} \quad H(x, Dv) \geq 0 \quad \text{in } \Omega$$

*in the viscosity sense. If we assume further that*



(1) there is a function  $\phi \in C^1(\Omega) \cap C(\bar{\Omega})$  such that  $\phi \leq u$  on  $\Omega$  and

$$\sup\{H(x, D\phi(x)) : x \in \omega, u \in \mathbb{R}\} < 0 \quad \forall \omega \subset\subset \Omega;$$

(2) the function  $p \rightarrow H(x, p)$  is convex on  $\mathbb{R}^n$  for each  $x \in \Omega$ ,

then  $u \leq v$  on  $\Omega$  provided that  $u - v$  is bounded in  $\Omega$  and  $u \leq v$  on  $\partial\Omega$  with  $\partial\Omega \neq \emptyset$ .

*Proof.* We use the same method as used to prove Lemma 1 and Theorem 1 in [Is] and Theorem 2.1 in [CEL]. We choose a smooth function  $\beta$  such that  $0 \leq \beta \leq 1$ ,  $\beta(0) = 1$  and  $\beta(x) = 0$  whenever  $\|x\| \geq 1$ . Let  $\beta_\epsilon(x) = \beta(x/\epsilon)$ .

Since  $u - v$  is assumed bounded, there exists a smooth function  $\phi$  such that both  $u - \phi$  and  $v - \phi$  are bounded. By a translation  $H(x, \partial u) \rightarrow H(x, \partial u + \partial \phi)$  we can assume that both  $u$  and  $v$  are bounded in  $\Omega$ .

To prove the first part of the lemma, we assume the contrary, namely,  $\exists x_0 \in \Omega$  such that  $u(x_0) - v(x_0) > 0$ . Let  $M = \max\{\|u\|, \|v\|\}$  and introduce  $\Phi: \Omega \times \Omega \rightarrow \mathbb{R}$  be given by

$$\Phi(x, x') = u(x) - v(x') + 3M\beta_\epsilon(x - x').$$

Because  $u - v \leq 0$  on  $\partial\Omega$  and  $\Phi \leq 2M$  if  $\|x - x'\| \geq \epsilon$ , one obtains that

$$\Phi(x_0, x_0) = u(x_0) - v(x_0) + 3M\beta_\epsilon(0) > 3M,$$

and

$$\Phi|_{\partial(\Omega \times \Omega)} < \Phi(x_0, x_0) \quad \text{if } \epsilon > 0 \text{ is sufficiently small.}$$

However, if  $\Omega$  is unbounded, the maximum of  $\Phi$  may not be reachable although it is bounded. Choose  $\delta > 0$  very small and there exists  $(x_1, x'_1)$  so that

$$\Phi(x_1, x'_1) > \sup_{\Omega \times \Omega} \Phi(x, x') - \delta.$$

We choose a smooth function  $\zeta: \Omega \times \Omega \rightarrow \mathbb{R}$  such that  $0 \leq \zeta \leq 1$ ,  $\zeta(x_1, x'_1) = 1$ ,  $\zeta(x, x') = 0$  whenever  $\|x - x_1\|^2 + \|x' - x'_1\|^2 > 1$  and  $|D\zeta| \leq 2$  in  $\Omega \times \Omega$ . Finally, we set

$$\Psi(x, x') = \Phi(x, x') + 2\delta\zeta(x, x').$$

Clearly,  $\Psi$  has a global maximum point  $(\bar{x}, \bar{x}') \in \Omega \times \Omega$ . Indeed,

$$\Psi(x_1, x'_1) = \Phi(x_1, x'_1) + 2\delta > \sup_{\Omega \times \Omega} \Phi + \delta$$

whereas

$$\limsup_{\|x\| + \|x'\| \rightarrow \infty} \Psi(x, x') \leq \sup_{\Omega \times \Omega} \Phi$$

and

$$\Psi(x, x')|_{\partial(\Omega \times \Omega)} < \sup_{\Omega \times \Omega} \Phi + \delta \quad \text{if } \delta > 0 \text{ is sufficiently small.}$$

Obviously, we have

$$\|\bar{x} - \bar{x}'\| < \epsilon.$$

Now  $\bar{x}$  is a maximum point of  $x \rightarrow u(x) - (v(\bar{x}') - 3M\beta_\epsilon(x - \bar{x}') - 2\delta\zeta(x, \bar{x}'))$  and thus, by assumption

$$(9.1) \quad H(\bar{x}, -3M\partial\beta_\epsilon(\bar{x} - \bar{x}') - 2\delta\partial_x\zeta(x, \bar{x}')) \leq f(\bar{x}).$$

Similarly,  $\bar{x}'$  is a minimum point of  $x' \rightarrow v(x') - (u(\bar{x}) + 3M\beta_\epsilon(x - \bar{x}') + 2\delta\zeta(x, \bar{x}'))$  and so

$$(9.2) \quad H(\bar{x}', -3M\partial\beta_\epsilon(\bar{x} - \bar{x}') + 2\delta\partial_{x'}\zeta(x, \bar{x}')) \geq 0.$$

Since  $f(\bar{x}) < 0$  and  $\epsilon$  as well as  $\delta$  can be chosen arbitrarily small, (9.1) contradicts to (9.2), which implies that  $u \leq v$  on  $\Omega$ . This completes the proof of the first part.

For the second part of the proof, we set

$$u_\theta(x) = \theta u(x) + (1 - \theta)\phi(x) \quad \text{for } x \in \bar{\Omega},$$

where  $\theta \in (0, 1)$ . By the conditions, we can choose function  $f \in C(\bar{\Omega})$  such that  $H(x, \partial\phi(x)) \leq f(x) < 0$  and  $f \leq u$  on  $\Omega$ . thus, we see that  $u_\theta \leq u$  on  $\Omega$  and  $u_\theta \in C(\bar{\Omega})$ . Note that  $p \rightarrow H(\cdot, p)$  is convex, a formal calculation reveals that

$$H(x, \partial u_\theta) \leq \theta H(x, \partial u) + (1 - \theta)H(x, \partial\phi) \leq f.$$

Applying the first part of the lemma and noting  $\theta$  can be arbitrarily close to 1, we complete the proof for the second part.  $\square$

*Proof of Theorem 9.1.* For each connected component  $\Omega$  of  $\Omega_{U', U}$  or  $\Omega_{U, U'}$ ,  $\exists$  a function  $g$  defined on  $\partial\Omega$  such that  $U_c|_{\partial\Omega} = U_{c'}|_{\partial\Omega} = g$ . Thus, both  $U_c$  and  $U_{c'}$  are the viscosity solution of the Dirichlet problem

$$\begin{cases} H(x, \partial u) = \alpha, & \text{in } \Omega, \\ u = g, & \text{on } \partial\Omega \end{cases}$$

where  $\alpha$  is valued as  $\alpha(c)$  and  $\alpha(c')$  respectively. Note,  $\partial\Omega \neq \emptyset$  and  $\Omega$  is not necessary bounded.

Clearly, the difference of  $U$  and  $U'$  is decomposed into a sum of bounded function and a linear function:  $U - U' = \Delta u + \langle c - c', x \rangle$ . Denote the level set of  $U - U'$  by

$$Z_{U, U'} = \{x \in \mathbb{R}^n : U(x) - U'(x) = 0\},$$

it is restricted a strip  $\{|\langle c - c', x \rangle| \leq D\}$  for certain positive number  $D > 0$ , i.e. both  $\Omega_{U', U}$  and  $\Omega_{U, U'}$  contain a unbounded connected component provided  $c \neq c'$ .

If  $\alpha(c) > \alpha(c')$ ,  $\Omega_{U', U}$  is connected, otherwise there would be a connected component contained in the strip  $\{|\langle c - c', x \rangle| \leq D\}$ . On this connected component  $U - U'$  is obviously bounded, thus the lemma is violated. Thus,  $\Omega_{U, U'}$  contains one unbounded connected component only. The possibility can not be excluded that  $\Omega_{U, U'}$  contains some bounded connected components.

For the case that  $\alpha(c) = \alpha(c') > \min \alpha$ , but  $c$  and  $c'$  are assumed not in one flat,  $\exists$   $0 < \lambda < 1$  such that

$$\alpha(c^*) < \alpha(c), \quad \text{where } c^* = \lambda c + (1 - \lambda)c'.$$

By the result in [FS], we know that there exists a  $C^1$  global sub-solution of the equation  $H(x, \partial u + c^*) = \alpha(c^*)$ ,  $\tilde{\phi}: \mathbb{T}^n \rightarrow \mathbb{R}$ , i.e.  $\phi(x) = \tilde{\phi}(x) + \langle x, c^* \rangle$  satisfies the condition  $H(x, \partial\phi) - \alpha(c^*) < 0$ . Since

$$U - \phi = \Delta u + (1 - \lambda)\langle c - c', x \rangle,$$

and

$$U' - \phi = \Delta u' - \lambda\langle c - c', x \rangle,$$

where both  $\Delta u$  and  $\Delta u'$  are periodic, thus  $U - \phi$  as well as  $U' - \phi$  can be set positive in the strip  $\{|\langle c - c', x \rangle| \leq D\}$  by adding a suitable constant to  $\phi$ . Since  $H$  is assumed convex in the action variable, by applying the second part of the lemma we find that both  $\Omega_{U, U'}$  and  $\Omega_{U', U}$  are simply connected and unbounded. This completes the proof for the second part.  $\square$

From this theorem we see that  $Z_{U,U'}$  consists of some pieces of  $(n-1)$ -dimensional “surface”, which may be not smooth. However, there exist some properties to describe their “normal” direction. The “derivative” set of  $U - U'$  at the point  $x$  is defined by

$$(9.3) \quad D_{U,U',x} = \{y - y' : y \in D^+U(x), y' \in D^+U'(x)\},$$

where  $D^+U$  denotes the sup-derivative of  $U$  (see Formula (A.3) in the appendix for the definition). Clearly,  $D_{U,U',x}$  is closed and convex. We define

$$(9.4) \quad \begin{aligned} S_{U,U',x}^{+,\delta} &= \{v \in \mathbb{R}^n : \langle y, v \rangle > \delta \|v\|, \forall y \in D_{U,U',x}\}, \\ S_{U,U',x}^{-,\delta} &= \{v \in \mathbb{R}^n : \langle y, v \rangle < -\delta \|v\|, \forall y \in D_{U,U',x}\}. \end{aligned}$$

If  $0 \notin D_{U,U',x}$ , both of them are not empty provided  $\delta > 0$  is suitably small.

Given a convex function  $\psi$ , for each direction  $e \in \mathbb{R}^n$  with  $\|e\| = 1$ , there exists a sub-derivative  $y_{\psi,e}$  of  $\psi$  at  $x$  such that

$$\psi(x + te) - \psi(x) = t \langle y_{\psi,e}, e \rangle + O(t^2) \quad t \geq 0.$$

Indeed, we have  $\langle y_{\psi,e}, e \rangle = \partial_e \psi(x)$  where  $\partial_e \psi$  is the derivative of  $\psi$  in the direction of  $e$ . Note each backward weak KAM solution has decomposition of a smooth function minus a convex function  $U = \phi - \psi$ . By definition, for  $x_1 = x + te$  with  $t > 0$  we have

$$(9.5) \quad \begin{aligned} U(x_1) - U(x) &= \langle y_e, x_1 - x \rangle + O(\|x_1 - x\|^2), \\ &\leq \langle y, x_1 - x \rangle + O(\|x_1 - x\|^2), \quad \forall y \in D^+U(x) \end{aligned}$$

where  $y_e = \partial \phi(x) - y_{\psi,e}$ . Similarly, we have

$$(9.6) \quad \begin{aligned} U'(x_1) - U'(x) &= \langle y'_e, x_1 - x \rangle + O(\|x_1 - x\|^2), \\ &\leq \langle y', x_1 - x \rangle + O(\|x_1 - x\|^2), \quad \forall y' \in D^+U'(x) \end{aligned}$$

where  $y'_e = \partial \phi'(x) - y_{\psi',e}$ . Therefore, if  $x \in Z_{U,U'}$  and  $x_1 - x \in S_{U,U',x}^{+,\delta}$ , we subtract (9.6) from (9.5) and obtain

$$\begin{aligned} U(x_1) - U'(x_1) &\geq \langle y_e - y'_e, x_1 - x \rangle - O(\|x_1 - x\|^2) \\ &\geq \delta \|x_1 - x\| - O(\|x_1 - x\|^2). \end{aligned}$$

Thus, there exists suitably small  $\epsilon > 0$  such that

$$(9.7) \quad U(x_1) - U'(x_1) > 0, \quad \forall x_1 - x \in S_{U,U',x}^{+,\delta} \cap B_\epsilon(x), \quad x \in Z_{U,U'},$$

where  $B_\epsilon$  is a ball in  $\mathbb{R}^n$ , centered at  $x$  with radius  $\epsilon$ . In the same way, we obtain

$$(9.8) \quad U(x_1) - U'(x_1) < 0, \quad \forall x_1 - x \in S_{U,U',x}^{-,\delta} \cap B_\epsilon(x), \quad x \in Z_{U,U'}.$$

**9.2. Modulus of continuity of weak KAM solutions.** on positive energy level of systems with two degrees of freedom.

In this section, our attention is concentrated on the systems with two degrees of freedom, i.e the configuration manifold is  $\mathbb{T}^2$ . We consider the level set of  $\alpha$ -function

$$\mathbf{C}_E = \{c \in H^1(M, \mathbb{R}) : \alpha(c) = E\}.$$

It is a closed and convex curve if  $E > \min \alpha$ . By adding a closed 1-form to the Lagrangian, we can assume that the  $\alpha$ -function reaches its minimal at zero cohomology. In the following, we shall assume this condition. In this case,  $\mathbf{C}_E$  encircles the origin.

**Proposition 9.1.** *Assume that  $\gamma$  and  $\gamma'$  are forward (backward)  $c$ -minimal and  $c'$ -minimal curves respectively, and  $\alpha(c) = \alpha(c')$ . Let  $\bar{\gamma}$  and  $\bar{\gamma}'$  denote their lift to the universal covering space, then  $\bar{\gamma}$  crosses  $\bar{\gamma}'$  at most once.*

*Proof.* We consider forward minimal case only, i.e. they are defined on  $[0, \infty)$ . The backward case is the same. Let us assume that they cross twice, i.e.  $\exists t_1 \neq t_2$  and  $s_1 \neq s_2$  such that  $\bar{\gamma}(t_i) = \bar{\gamma}'(s_i)$  for  $i = 1, 2$ . Clearly,  $\theta_i = \dot{\gamma}(t_i) - \dot{\gamma}'(s_i) \neq 0$ . By a fundamental lemma of Mather in [M91] we obtain that some  $\epsilon > 0$  and  $C > 0$  exist depending on the Lagrangian only such that

$$A(\gamma|_{[t_i-\epsilon, t_i+\epsilon]}) + A(\gamma'|_{[s_i-\epsilon, s_i+\epsilon]}) - A(a_i) - A(b_i) \geq C\epsilon\|\theta_i\|^2,$$

where  $a_i: [-\epsilon, \epsilon] \rightarrow \mathbb{R}^2$  is a minimal curve of  $L$  joining  $\bar{\gamma}'(s_1 - \epsilon)$  to  $\bar{\gamma}(t_i + \epsilon)$ , i.e.  $a_i(-\epsilon) = \bar{\gamma}'(s_1 - \epsilon)$ ,  $a_i(\epsilon) = \bar{\gamma}(t_i + \epsilon)$  and

$$A(a_i) = \int_{-\epsilon}^{\epsilon} L(a(t), \dot{a}(t)) dt = \inf_{\substack{\zeta(-\epsilon)=a(-\epsilon) \\ \zeta(\epsilon)=a(\epsilon)}} \int_{-\epsilon}^{\epsilon} L(\zeta(t), \dot{\zeta}(t)) dt,$$

$b_i: [-\epsilon, \epsilon] \rightarrow \mathbb{R}^2$  is a minimal curve of  $L$  joining  $\bar{\gamma}(s_1 - \epsilon)$  to  $\bar{\gamma}'(t_i + \epsilon)$  and  $A(b_i)$  is defined in the same way as for  $A(a_i)$ . Let  $\Delta s = s_2 - s_1$ ,  $\Delta t = t_2 - t_1$ , we define two curves

$$\xi(t) = \begin{cases} \gamma(t), & t \in [0, t_1 - \epsilon], \\ b_1(t - t_1), & t - t_1 \in [-\epsilon, \epsilon], \\ \gamma'(t - t_1), & t - t_1 \in [s_1 + \epsilon, s_2 - \epsilon], \\ a_2(t - t_1 + \Delta s), & t - t_1 + \Delta s \in [-\epsilon, \epsilon], \\ \gamma(t - t_1 + \Delta s), & t - t_1 + \Delta s \in [t_2 + \epsilon, \infty), \end{cases}$$

and

$$\xi'(t) = \begin{cases} \gamma'(t), & t \in [0, s_1 - \epsilon], \\ a_1(t - s_1), & t - s_1 \in [-\epsilon, \epsilon], \\ \gamma(t - s_1), & t - s_1 \in [t_1 + \epsilon, t_2 - \epsilon], \\ b_2(t - s_1 + \Delta t), & t - s_1 + \Delta t \in [-\epsilon, \epsilon], \\ \gamma'(t - s_1 + \Delta t), & t - s_1 + \Delta t \in [s_2 + \epsilon, \infty). \end{cases}$$

By the construction, we have  $\gamma(t_1 - \epsilon) = \gamma'(t_1 - \epsilon)$ ,  $\gamma(t_2 + \epsilon) = \gamma'(t_1 + \Delta s + \epsilon)$ ,  $\gamma'(s_1 - \epsilon) = \gamma''(s_1 - \epsilon)$ ,  $\gamma'(s_2 + \epsilon) = \gamma''(s_1 + \Delta t + \epsilon)$  and

$$\begin{aligned} & [A_c(\gamma|_{[t_1-\epsilon, t_2+\epsilon]})] + [A_{c'}(\gamma'|_{[s_1-\epsilon, s_2+\epsilon]})] - [A_c(\xi|_{[t_1-\epsilon, t_1+\Delta s+\epsilon]})] \\ & - [A_{c'}(\xi'|_{[s_1-\epsilon, s_1+\Delta t+\epsilon]})] \\ & = A(\gamma|_{[t_1-\epsilon, t_1+\epsilon]}) + A(\gamma'|_{[s_1-\epsilon, s_1+\epsilon]}) - A(a_1) - A(b_1) \\ & \quad + A(\gamma|_{[t_2-\epsilon, t_2+\epsilon]}) + A(\gamma'|_{[s_2-\epsilon, s_2+\epsilon]}) - A(a_2) - A(b_2) \\ & \geq C\epsilon(\|\theta_1\|^2 + \|\theta_2\|^2). \end{aligned}$$

This contradicts that fact that  $\gamma$  and  $\gamma'$  are  $c$ - and  $c'$ -minimal curves respectively. The absurdity verifies the fact that  $\bar{\gamma}$  crosses  $\bar{\gamma}'$  at most once.  $\square$

**Corollary 9.1.** *Assume that both  $\gamma$  and  $\gamma'$  are forward (backward)  $c$ -minimal curves,  $d\gamma$  and  $d\gamma'$  share the same  $\omega$ - or  $\alpha$ -limit set. Let  $\bar{\gamma}$  and  $\bar{\gamma}'$  denote their lift to the universal covering space, then they do not cross everywhere.*

**Lemma 9.2.** *Assume the autonomous Lagrangian is defined on two-torus. For each  $c \in \mathbf{C}_E$  with  $E > \min \alpha$ , the rotation vectors of all orbits in  $\tilde{\mathcal{M}}(c)$  have the same direction, i.e. if  $d\gamma_1, d\gamma_2 \in \tilde{\mathcal{M}}(c)$ , then*

$$\langle \omega(\gamma_1), \omega(\gamma_2) \rangle = \|\omega(\gamma_2)\| \|\omega(\gamma_2)\| > 0.$$

*Proof.* We claim that for each orbit  $d\gamma: \mathbb{T}^n \rightarrow TM$  in the support of the minimal measure, we have  $\langle c - c^*, [\gamma] \rangle \neq 0$ , where  $c^*$  is the minimum point of the  $\alpha$ -function, which is defaulted to be zero here. If it was not true, there would be a sequence of time  $T_i \rightarrow \infty$  as  $i \rightarrow \infty$  such that  $\lim_{i \rightarrow \infty} \gamma(T_i) = \gamma(0)$  and

$$\begin{aligned} \frac{1}{T_i} \int_0^{T_i} L(d\gamma(t)) dt &= \frac{1}{T_i} \int_0^{T_i} (L(d\gamma(t)) - \langle c, \dot{\gamma}(t) \rangle) dt \\ &= -E < -\min \alpha, \end{aligned}$$

but it is absurd. As an immediate consequence,  $[\gamma] \neq 0$  holds for each  $d\gamma$  in the support of the minimal measure.

Next, we see that the rotation vector of one orbit in the Mather set is co-linear with the rotation vector of any other orbits in the set. Otherwise, they would intersect each other for infinitely many times. But it contradicts the Lipschitz graph property of Aubry set.

Finally, we show that all of the rotation vectors have the same direction. Let us assume the contrary, i.e. there exist two orbits  $\gamma_j: \mathbb{R} \rightarrow M$  ( $j = 1, 2$ ) such that

$$[\gamma_1] = -[\gamma_2]$$

and

$$\int_0^{T_{i_j}} L(d\gamma_i(t)) dt - \langle c, [\gamma_i] \rangle + T_{i_j} E = 0, \quad j = 1, 2,$$

where  $T_{i_j} \rightarrow \infty$  as  $i_j \rightarrow \infty$  such that  $\lim_{i_j \rightarrow \infty} \gamma(T_{i_j}) = \gamma_j(0)$ . As  $\langle c, [\gamma_i] \rangle \neq 0$  for  $i = 1, 2$  and  $-\gamma_1 = \gamma_2$ , we can assume that  $\langle c, [\gamma_1] \rangle < 0$ . As  $\mathbf{C}_E$  is a closed and convex curve encircling the origin, there exist  $c' \in \mathbf{C}_E$  and  $\lambda > 0$  such that  $c' = -\lambda c$ . By computing the  $c'$ -average action along  $\gamma_1$ , we find

$$\begin{aligned} &\frac{1}{T_{i_1}} \int_0^{T_{i_1}} (L(d\gamma(t)) - \langle c', [\gamma_1] \rangle + E) dt \\ &= \frac{1}{T_{i_1}} \int_0^{T_{i_1}} (L(d\gamma(t)) - \langle c, [\gamma_1] \rangle + E + \langle (1 + \lambda)c, [\gamma_1] \rangle) dt \\ &\rightarrow \langle (1 + \lambda)c, [\gamma_1] \rangle < 0. \end{aligned}$$

The contradiction completes the proof.  $\square$

Each energy level has a natural fibration. The fiber over a point  $x$  is denoted by

$$\mathbf{Y}_{x,E} = \{y : (x, y) \in H^{-1}(E)\}.$$

If  $E > \min \alpha$ ,  $\mathbf{Y}_{x,E}$  is a smooth, convex and closed curve for each  $x \in M$ . Indeed, because of Proposition A.2, there is at least one backward semi-static curve  $\gamma_c^-$  for each  $c \in \mathbf{C}_E$  such that  $\gamma_c^-(0) = x$ . Clearly,  $\partial_x L(\gamma_c^-(0), \dot{\gamma}_c^-(0)) \in \mathbf{Y}_{x,E}$ . As  $E > \min \alpha$ ,  $\mathbf{C}_E$  contains at least three vertex, namely, there are at least three cohomology classes in  $\mathbf{C}_E$  such that the semi-static curves are different. Consequently,  $\mathbf{Y}_{x,E}$  is a closed curve. The smoothness and the convexity are obvious. Let

$$\mathbf{V}_{x,E} = \{v = \partial_y H(x, y) : y \in \mathbf{Y}_{x,E}\},$$

it is a smooth, convex and closed curve in  $\mathbb{R}^2$  encircling the origin.

Given three vectors  $v_1, v_2, v_3 \in \mathbf{V}_{x,E} \subset \mathbb{R}^2$ , we say that they are in clock-wise order, denoted by  $v_1 \prec v_2 \prec v_3$  if the points  $\frac{v_1}{\|v_1\|}, \frac{v_2}{\|v_2\|}, \frac{v_3}{\|v_3\|}$  are in clock-wise order on the unit circle. Via the one-to-one correspondence  $v \rightarrow y = \partial_x L(x, v)$ ,  $v_1, v_2, v_3$  in  $\mathbf{V}_{x,E}$  uniquely determine three points  $y_1, y_2, y_3$  in  $\mathbf{Y}_{x,E}$ . As  $L$  is strictly positive definite

in the speed, we have  $y_1 - y_x \prec y_2 - y_x \prec y_3 - y_x$  where  $y_x$  is the minimal point of  $H(x, y)$  with fixed  $x$ .

For each  $y \in \mathbf{Y}_{x,E}$ , there is a unique smooth curve  $\gamma(t, x, y): \mathbb{R} \rightarrow M$  such that  $\gamma(0, x, y) = x$ ,  $\partial_{\dot{x}}L(x, \dot{\gamma}(0, x, y)) = y$ . Indeed,  $(\gamma(t), \dot{\gamma}(t))$  is a trajectory of the Lagrange flow. However, it is not necessary that each of these curve is semi-static. Let

$$\mathbf{Y}_{x,E}^- = \{y \in \mathbf{Y}_{x,E} : \gamma(\cdot, x, y)|_{t \in \mathbb{R}^-} = \gamma_c^-(\cdot, x) \text{ for some } c \in \mathbf{C}_E\},$$

namely, each point  $y \in \mathbf{Y}_{x,E}^-$  determines a backward semi-static curve  $\gamma(\cdot, x, y)$  for certain cohomology class  $c$  such that  $\gamma(0, x, y) = x$  and  $\partial_{\dot{x}}L(\gamma(0, x, y), \dot{\gamma}(0, x, y)) = y$ . It approaches to certain Aubry set in the following sense

$$\alpha(d\gamma(\cdot, x, y)) \subseteq \tilde{\mathcal{A}}(c).$$

Since the configuration space is two-dimensional, each orbit in  $\alpha(d\gamma(\cdot, x, y)) \cap \tilde{\mathcal{A}}(c)$  has the same rotation vector, denoted by  $\omega(x, y)$ . Because of upper semi-continuity of backward semi-static curves on cohomology classes, the set  $\mathbf{Y}_{x,E}^-$  is closed in  $\mathbf{Y}_{x,E}$ .

As the complementary set  $\mathbf{Y}_{x,E} \setminus \mathbf{Y}_{x,E}^-$  is composed of open intervals in  $\mathbf{Y}_{x,E}$ , an equivalence relation  $\sim$  in  $\mathbf{Y}_{x,E}$  is introduced such that  $y \sim y'$  if  $y$  and  $y'$  are the two boundary points of an open interval in  $\mathbf{Y}_{x,E} \setminus \mathbf{Y}_{x,E}^-$ .

**Proposition 9.2.** *If  $E > \min \alpha$ ,  $y \sim y'$  in  $\mathbf{Y}_{x,E}^-$ , then  $\omega(x, y)$  and  $\omega(x, y')$  have the same direction, i.e.*

$$\langle \omega(x, y), \omega(x, y') \rangle = \|\omega(x, y)\| \|\omega(x, y')\|.$$

*If  $c$  and  $c'$  are the cohomology classes such that  $\gamma(\cdot, x, y)$ ,  $\gamma(\cdot, x, y')$  are the  $c$ -,  $c'$ -semi static respectively, then they stay in the one flat of the  $\alpha$ -function.*

*Proof.* Let us assume the contrary, namely, the direction of  $\omega(x, y)$  is different from that of  $\omega(x, y')$ . Under such assumption, the curves  $\gamma(\cdot, x, y)|_{t \in \mathbb{R}^-}$  and  $\gamma(\cdot, x, y')|_{t \in \mathbb{R}^-}$  can not be semi-static for the same cohomology class, it is guaranteed by Lemma 9.2.

Denoted by  $c$  and  $c'$  the cohomology classes such that  $\gamma(\cdot, x, y)|_{t \in \mathbb{R}^-}$  is  $c$ -semi-static and  $\gamma(\cdot, x, y')|_{t \in \mathbb{R}^-}$  is  $c'$ -semi-static. We claim that  $c$  and  $c'$  are not contained in one flat of the  $\alpha$ -function under the assumption. Indeed, by the result of [Ms],  $\mathcal{A}(c_1) = \mathcal{A}(c_2)$  whenever both  $c_1$  and  $c_2$  are in the relative interior of the flat,  $\mathcal{A}(c_1) \supset \mathcal{A}(c_2)$  if  $c_1$  is on the boundary of the flat while  $c_2$  is in the relative interior. Choose  $c^*$  from the relative interior of the flat where  $c, c'$  are, then we have  $\mathcal{A}(c) \supseteq \mathcal{A}(c^*)$  and  $\mathcal{A}(c') \supseteq \mathcal{A}(c^*)$ . But this places us in a dilemma: if the direction of the rotation vector for  $\tilde{\mathcal{A}}(c^*)$  is the same as that for  $\tilde{\mathcal{A}}(c)$ , then it is different from that for  $\tilde{\mathcal{A}}(c')$ , and vice verse.

**Lemma 9.3.** *If an autonomous Lagrangian is defined on two dimensional torus, then, any flat of the  $\alpha$ -function is disjoint to any other flat.*

*Proof.* If there are two flats sharing a common boundary point, the derivative of the  $\alpha$ -function at one flat is not colinear with the derivative at another one. Therefore, for the cohomology class at the common boundary point, the Mather set contains orbits with non colinear rotation vectors. However, it is impossible when the configuration space is a two dimensional torus.  $\square$

According to this lemma, we see that  $\mathbf{C}_E$  is a closed and convex curve with infinitely many edges whenever  $E > \min \alpha$ . As  $\mathbf{C}_E$  is a convex and closed curve encircling the

origin in this case, other two cohomology classes  $c_1, c_2 \in \mathbf{C}_E$  exist such that

$$c_1 \prec c \prec c_2 \prec c' \prec c_1,$$

and no flat contains any two of these cohomology classes. Let  $\omega_1, \omega, \omega_2, \omega'$  be the rotation vectors conjugate to  $c_1, c, c_2, c'$  via Fenchel-Legendre transformation  $\mathcal{L}_\beta^{-1}$  respectively

$$\omega \in \mathcal{L}_\beta^{-1}(c) \Rightarrow \langle \omega, c \rangle = \alpha(c) + \beta(\omega),$$

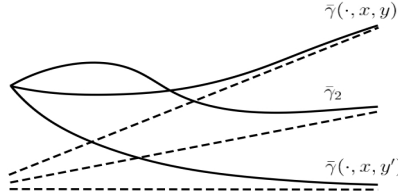
then we have

$$\omega_1 \prec \omega \prec \omega_2 \prec \omega' \prec \omega_1.$$

By choose suitable  $c_1$  and  $c_2$ , we assume that any two of these rotation vectors are not colinear. Guaranteed by Proposition A.2, there exists a backward  $c_i$ -semi static curve  $\gamma_i^-$  ( $i = 1, 2$ ) such that  $\gamma_i^-(0) = x$ . Denoted  $y_i = \partial_{\dot{x}} L(\gamma_i^-(0), \dot{\gamma}_i^-(0))$ , we claim that

$$y_1 - y_x \prec y - y_x \prec y_2 - y_x \prec y' - y_x \prec y_1 - y_x.$$

Indeed, if  $y_2 - y_x \prec y - y_x \prec y' - y_x$ , we would have  $\dot{\gamma}_2^-(0) \prec \dot{\gamma}(0, x, y) \prec \dot{\gamma}(0, x, y')$ , because  $\partial_{\dot{x}\dot{x}}^2 L$  is positive definite. Let  $\bar{\gamma}(\cdot, x, y)$ ,  $\bar{\gamma}(\cdot, x, y')$  and  $\bar{\gamma}_2$  denote the lift of the curves  $\gamma(\cdot, x, y)$ ,  $\bar{\gamma}(\cdot, x, y')$  and  $\gamma_2$  to the universal covering space  $\mathbb{R}^2$  respectively such that  $\bar{\gamma}(0, x, y) = \bar{\gamma}(0, x, y') = \gamma_2(0)$ . In this case,  $\bar{\gamma}_2$  either cross  $\bar{\gamma}(\cdot, x, y)$  or cross  $\bar{\gamma}(\cdot, x, y')$  at another point, because of the order  $\omega \prec \omega_2 \prec \omega'$ , see Figure 2. But this violates the property that all these three curves are backward semi-static. This contradiction verifies our claim. However, the induced conclusion  $y_1 - y_x \prec y - y_x \prec$



$y_2 - y_x \prec y' - y_x \prec y_1 - y_x$  contradicts the hypotheses that  $y \sim y'$ . It implies that  $\omega(x, y)$  and  $\omega(x, y')$  have the same direction. Consequently,  $c$  and  $c'$  are in one flat of the  $\alpha$ -function.  $\square$

Let us consider all of those Aubry sets  $\{\mathcal{A}_c\}$  such that the rotation vector  $\omega(\mu_c)$  is non-resonant. In this case, each  $c$ -minimal measure is uniquely ergodic and the lift of the Aubry set to any finite covering  $k\mathbb{T}^2 = \{(x_1, x_2) \in \mathbb{R}^2 : x_i \bmod k_i\}$  still consists of only one Aubry class. Let  $u_{c,\infty}^-$  be the (backward) globally elementary weak KAM solution for  $\mathcal{A}_c$ . We consider the set of globally elementary weak KAM solutions

$$\mathcal{U}_E^\pm = \{U_{c,\infty}^\pm = u_{c,\infty}^\pm + \langle c, x \rangle : \forall c \in \mathbf{C}_E, \omega(\mu_c) \text{ is non-resonant}\},$$

and let

$$\mathcal{U}_{E,K}^\pm = \{U_{c,i}^\pm|_{B_K} : U_{c,i}^\pm \in \mathcal{U}_E^\pm\} \subset C^0(B_K).$$

Let  $\bar{\mathcal{U}}_{E,K}^\pm$  be the closure of  $\mathcal{U}_{E,K}^\pm$  in  $C^0$ -topology.

**Theorem 9.2.** *Let  $M = \mathbb{T}^2$  and  $E > \min \alpha$ , then for each finite  $K > 0$ , the set  $\bar{\mathcal{U}}_{E,K}^\pm$  can be parameterized by some parameter  $\sigma$  defined in a closed interval such that  $\sigma \rightarrow U_\sigma^\pm$  is  $\frac{1}{3}$ -Hölder continuous in  $C^0$ -topology.*

*Proof.* As the first step, we consider the set  $\mathcal{U}_{E,K}^\pm$ . Let  $\omega_i$  denote the rotation vector of the  $c_i$ -minimal measure,  $U_i^-$  be the globally weak KAM solution,  $\gamma_{c_i,x}^-: (-\infty, 0] \rightarrow M$  be the backward semi-static curve for the cohomology classes  $c_i$  such that  $\gamma_{c_i,x}^-(0) = x$ ,  $\partial_x L(\gamma_{c_i,x}^-(0), \dot{\gamma}_{c_i,x}^-(0)) + c_i = \partial_x U_i^-(x)$  whenever  $U_i^-$  is differentiable at  $x$ . Denoted by  $v_i = \dot{\gamma}_{c_i,x}^-(0)$  and consider them as vectors, we see that  $v_{i_1} \prec v_{i_2} \prec v_{i_3}$  provided  $\omega_{i_1} \prec \omega_{i_2} \prec \omega_{i_3}$ , see the proof of Proposition 9.2.

For each point  $x \in \mathbb{T}^2$  and each  $c_i$ , there exists at least one backward semi-static curve  $\gamma_{c_i,x}^-: (-\infty, 0] \rightarrow \mathbb{T}^2$  such that  $\gamma_{c_i,x}^-(0) = x$  and  $\pi\alpha(d\gamma_{c_i}^-) \subset \mathcal{M}(c_i)$ . Let  $y_i = \partial_x L(v_i, x)$ , then  $y_i \in \mathbf{Y}_{x,E}^-$  for  $i = i_1, i_2, i_3$  and  $y_{i_1} - y_x \prec y_{i_2} - y_x \prec y_{i_3} - y_x$  provided  $v_{i_1} \prec v_{i_2} \prec v_{i_3}$ .

Given  $x \in \mathbb{T}^2$ , there might be two backward  $c$ -semi-static curve originating from this point, denoted by  $\gamma_c^-$  and  $\gamma_c'^-$  with  $v = \dot{\gamma}_c^-(0) \neq v' = \dot{\gamma}_c'^-(0)$ . In this case, the weak KAM solution is not differentiable at this point, there are at least two points  $y_c, y_c' \in \mathbf{Y}_{x,E}^-$  such that  $y_c = \partial_x L(v_c, x)$  and  $y_c' = \partial_x L(v_c', x)$ . The closed curve  $\mathbf{Y}_{x,E}$  is divided into two segments of arc, both have  $y_c$  and  $y_c'$  as their end points,  $I_{x,\omega} \cup I'_{x,\omega} = \mathbf{Y}_{x,E}$ , where  $\omega$  is the associated rotation direction. We claim that either  $I'_{x,\omega}$  or  $I_{x,\omega}$  is forbidden for any  $\omega' \neq \omega$  in the following sense: for each  $y$  in this segment and for each  $\omega' \neq \omega$ ,  $(x, y)$  does not determine any backward  $c'$ -semi-static orbit which approaches to the support of certain minimal measure with rotation direction  $\omega'$ . If not, there would be a backward  $c'$ -semi-static curve  $\gamma_{c'}^-$  as well as a backward  $c^*$ -semi-static curve  $\gamma_{c^*}^-$ , originating from  $x$  with associated rotation direction  $\omega', \omega^*$  satisfying the order condition  $\omega' \prec \omega \prec \omega^* \prec \omega$ . In this case, either  $\gamma_c^-$  or  $\gamma_c'^-$  would cross  $\gamma_{c'}^-$  or  $\gamma_{c^*}^-$  somewhere. But it is absurd.

Let  $y_i \in \mathbf{Y}_{x,E}^-$  such that  $y_i = \partial_x L(\dot{\gamma}_{c_i,x}^-(0), x)$ . By the notation we used before,  $y_i \in D^+U_i^-(x)$ . As  $y_i \prec y_j \prec y_k$  holds for each  $y_i \in D^+U_i^-(x)$ ,  $y_j \in D^+U_j^-(x)$  and  $y_k \in D^+U_k^-(x)$ , we use the notation

$$(9.9) \quad D^+U_i^-(x) \prec D^+U_j^-(x) \prec D^+U_k^-(x)$$

to imply this relation. Different  $x$  determines different  $y_i$  (might be multi-valued). However, the order property is independent of the position  $x$ , i.e.  $y_{i_1} \prec y_{i_2} \prec y_{i_3}$  holds everywhere once it holds at some point  $x \in M$ , namely, the relation (9.9) is independent of  $x$ .

By choosing suitable constants we assume that  $U_i^-(x_0) = 0$  for each  $i$ , where the point  $x_0$  is pretty far away from the origin. Denoted by

$$Z_{i,j}^- = \{x \in \mathbb{R}^2 : U_j^- - U_i^- = 0\}.$$

As these rotation direction are all different, any two of the cohomology classes are clearly not in the same flat. In virtue of Theorem 9.1, each  $Z_{i,j}^-$  is a continuous curve not containing a segment encircling an bounded region. Thus it extends to infinity.

**Lemma 9.4.** *Assume that  $\alpha(c_i) = \alpha(c_j) = \alpha(c_k)$ , the rotation vectors  $\omega_i, \omega_j$  and  $\omega_k$  are non-resonant and  $\omega_i \neq \omega_j \neq \omega_k$ , then  $Z_{i,j}^-$  intersects  $Z_{i,k}^-$  only at  $x = x_0$ .*

*Proof.* By Theorem A.1, the vertex of  $D^+U_i^-(x)$  must be on  $\mathbf{Y}_{x,E}^-$ . Thus, one deduces from the assumption  $\omega_i \neq \omega_j \neq \omega_k$  that the “derivative” sets (see (9.3) for the definition)  $D_{U_i^-, U_j^-, x}$ ,  $D_{U_i^-, U_k^-, x}$  and  $D_{U_j^-, U_k^-, x}$  are disjoint to each other for each  $x \in M$ , non of them contains the origin. If  $\omega_i \prec \omega_j \prec \omega_k$ , then  $y_j - y_i \prec y_k - y_i \prec y_k - y_j$ ,



namely

$$(9.10) \quad D_{U_i^-, U_j^-, x} \prec D_{U_i^-, U_k^-, x} \prec D_{U_j^-, U_k^-, x}.$$

Consequently, each intersection point of  $Z_{i,j}^-$  with  $Z_{i,k}^-$  is isolated. To see it, let us note that  $0 \notin \text{co}D_{U_i^-, U_j^-, x}$  holds for any  $x \in M$  and  $i \neq j$ . If  $Z_{i,j}^-$  intersects  $Z_{i,k}^-$  at some point  $x$ , then  $U_i^-(x) = U_j^-(x) = U_k^-(x)$ . Let  $x' \in Z_{i,j}^-$  be a point close to  $x$ , we obtain from the formula (9.7) and (9.8) that  $x' - x$  is almost orthogonal to certain vector  $y \in D_{U_i^-, U_j^-, x}$ :  $\langle x' - x, y \rangle \leq \delta \|x' - x\| \|y\|$  holds for sufficiently small  $\delta > 0$ . Therefore, we obtain from (9.10) that

$$Z_{i,j}^- \cap B_\epsilon \subset B_\epsilon \setminus (S_{U_i^-, U_j^-, x}^{+, \delta} \cup S_{U_i^-, U_j^-, x}^{-, \delta}),$$

$$Z_{i,k}^- \cap B_\epsilon \subset S_{U_i^-, U_j^-, x}^{+, \delta} \cup S_{U_i^-, U_j^-, x}^{-, \delta}$$

provided  $\delta > 0$  is sufficiently small. It implies that  $x$  is the only point in  $B_\epsilon(x)$  where  $Z_{i,j}^-$  intersects  $Z_{i,k}^-$  and the intersection is “topologically transversal”.

Therefore, for an intersection point  $x$  of  $\Gamma_{i,j}$  with  $\Gamma_{i,k}$ , it makes sense to find another intersection point  $x'$  next to  $x$  if there are more than one intersection point. In this case,  $\Gamma_{j,k}$  also passes through the points  $x$  and  $x'$ . Note that the intersection of these curves is always topologically transversal. If the “gradient” sets of these curves at  $x$  are in clock-wise order

$$D_{U_i^-, U_j^-, x} \prec D_{U_i^-, U_k^-, x} \prec D_{U_j^-, U_k^-, x},$$

then at  $x'$  they would be in anti clock-wise order (see Figure 6)

$$D_{U_i^-, U_j^-, x'} \succ D_{U_i^-, U_k^-, x'} \succ D_{U_j^-, U_k^-, x'}.$$

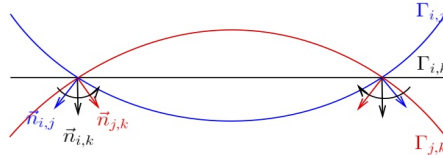


FIGURE 6

However, the order property (9.9) is shown independent of  $x$ . The contradiction verifies the fact: they intersect each other only at  $x = x_0$ .  $\square$

Given level sets  $Z_{j_\ell, i_\ell}^-$ ,  $\ell = 1, 2, \dots$ . If they intersect each other only at the origin, and at this point the “gradient” of  $U_{j_\ell}^- - U_{i_\ell}^-$  are in the order

$$D_{U_{j_1}^-, U_{i_1}^-, x_0} \prec D_{U_{j_2}^-, U_{i_2}^-, x_0} \prec D_{U_{j_3}^-, U_{i_3}^-, x_0} \prec \dots,$$

then we say these curves are in the order

$$Z_{j_1, i_1}^- \prec Z_{j_2, i_2}^- \prec Z_{j_3, i_3}^- \prec \dots.$$

Given two points  $y_\ell, y_k \in \mathbf{Y}_{x_0, E}^-$ , we choose other two points  $y_0, y_1 \in \mathbf{Y}_{x_0, E}^-$  such that they are in the order  $y_0 \prec y_\ell \prec y_k \prec y_1$ . We assume that each of these four points determines an backward minimal curve. Thus, we have the order:

$$y_\ell - y_0 \prec y_k - y_0 \prec y_k - y_\ell \prec y_1 - y_\ell \prec y_1 - y_k.$$

By applying Lemma 9.4 to the order  $y_\ell - y_0 \prec y_k - y_0 \prec y_k - y_\ell$  as well as to the order  $y_k - y_\ell \prec y_1 - y_\ell \prec y_1 - y_k$ , we see that the order property holds

$$Z_{\ell,0}^- \prec Z_{k,0}^- \prec Z_{k,\ell}^- \prec Z_{1,\ell}^- \prec Z_{1,k}^-.$$

Let us look this order property from another point of view.  $\Gamma_{k,\ell}$  lies in the sector-shaped region bounded by the lines  $\Gamma_{\ell,0}$  and  $\Gamma_{1,k}$  containing the curves  $\Gamma_{k,0}$  and  $\Gamma_{1,\ell}$ , see Figure 7.

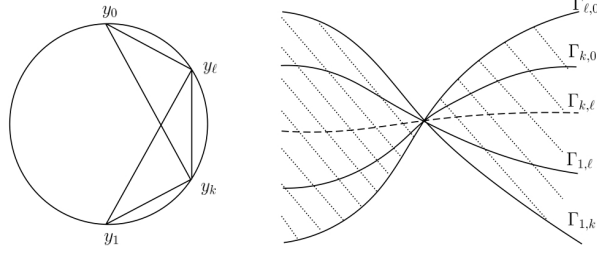


FIGURE 7

For any points  $y_i, y_j \in \mathbf{Y}_{x_0,E}^-$  between  $y_\ell$  and  $y_k$  in the following sense

$$y_0 \prec y_\ell \prec y_i \prec y_j \prec y_k \prec y_1,$$

then we obtain the order property by applying the same argument

$$Z_{\ell,0}^- \prec Z_{i,0}^- \prec Z_{j,i}^- \prec Z_{1,j}^- \prec Z_{1,k}^-,$$

namely, the formula (9.10) verifies the fact that  $Z_{j,i}^-$  lies in the sector-shaped region bounded by the lines  $Z_{\ell,0}^-$  and  $Z_{1,k}^-$  and containing the curve  $Z_{k,\ell}^-$ , provided  $U_i^-$  and  $U_j^-$  correspond to different rotation direction. Therefore, by choosing point  $x_0$  suitably far away from the origin,  $Z_{j,i}^- \cap B_K = \emptyset$ .

Therefore, this subset of  $\mathcal{U}_{E,K}^-$  is totally ordered provided each of them is valued zero at the point  $x_0$ :  $U_i^-(x_0) = 0$ . For any two function  $U_i^-, U_j^-$  in this set we have either  $U_i^-(x) - U_j^-(x) \geq 0$  for all  $x \in B_K$ , or alternatively,  $U_i^-(x) - U_j^-(x) \leq 0$  for all  $x \in B_K$ . Therefore, this subset of  $\mathcal{U}_{E,K}$  can be parameterized by “volume” in the following way, fix any  $i = i(0)$ , other  $i(\sigma)$  is defined such that

$$\sigma = \int_{B_K} (U_{i(\sigma)}^- - U_{i(0)}^-) dx$$

where the integration is in the sense of Lebesgue. The order property guarantees that the inverse of the map  $i \rightarrow \sigma$  is well-defined. Thus, this subset of  $\mathcal{U}_{E,K}^-$  can be parameterized by volume  $\sigma$ . Notice that each non-resonant rotation vector corresponds to unique first cohomology class  $c$ , one obtains the dependence  $c = c(\sigma)$  if  $\omega(\mu_c)$  is non-resonant.

With respect to  $\sigma$ , the weak KAM solutions have  $\frac{1}{3}$ -Hölder regularity in  $C^0$ -topology. The norm is given

$$\|U_\sigma^- - U_{\sigma'}^-\| = \max_{x \in B_K} |U_\sigma^-(x) - U_{\sigma'}^-(x)|.$$

Note these functions are all Lipschitz with common constant  $d$ . Given  $U_\sigma^- > U_{\sigma'}^-$ , the region

$$D_{\sigma,\sigma'} = \{(x, z) \in \Omega \times \mathbb{R} : U_\sigma^- \leq z \leq U_{\sigma'}^-\}$$

contains at least two quarters of a cone, its height is  $\frac{1}{2}\|U_\sigma^- - U_{\sigma'}^-\|$ , and the radius of its bottom disc is at least  $\frac{1}{2d}\|U_\sigma^- - U_{\sigma'}^-\|$ , i.e.

$$|\sigma - \sigma'| = \left| \int_{B_K} (U_\sigma^- - U_{\sigma'}^-) d\mu \right| \geq \frac{\pi}{6d^2} \|U_\sigma^- - U_{\sigma'}^-\|^3,$$

from which one obtains the  $\frac{1}{3}$ -Hölder regularity immediately.

Since  $\mathbf{Y}_{x_0, E}^-$  can be covered by several sets  $\{y_\theta : \theta \in \Theta_i\}$ , we obtain the regularity of the whole  $\mathcal{U}_{E, K}^\pm$ . The regularity extends naturally to the whole  $\bar{\mathcal{U}}_{E, K}^\pm$ .  $\square$

As all weak KAM solutions are Lipschitz functions with bounded Lipschitz constant, each function in  $\bar{\mathcal{U}}_{E, K}^\pm \setminus \mathcal{U}_{E, K}^\pm$  is also weak KAM solutions for  $c \in \mathbf{C}_E$  such that  $\omega(\mu_c)$  is resonant. The circle  $\mathbf{C}_E$  contains countably many edges (flat)  $\mathbb{F}_{E, i}$  in general. For  $c \in \mathbb{F}_{E, i}$ , the  $c$ -minimal measure is supported on periodic orbit. Let  $c_{i, \ell}$ ,  $c_{i, r}$  be the endpoint of  $\mathbb{F}_{E, i}$ , and let  $\{c_{k, \ell}\}$ ,  $\{c_{k, r}\}$  be two sequences such that both  $\omega(c_{k, \ell})$  and  $\omega(c_{k, r})$  are non-resonant,  $c_{k, \ell} \rightarrow c_{i, \ell}$  and  $c_{k, r} \rightarrow c_{i, r}$  as  $k \rightarrow \infty$ , one obtains weak KAM solutions  $U_{c_{i, \ell}} = \lim_{k \rightarrow \infty} U_{c_{k, \ell}}^-$  and  $U_{c_{i, r}} = \lim_{k \rightarrow \infty} U_{c_{k, r}}^-$ . The subscript  $\ell$  and  $r$  indicate the left and the right respectively. If  $U_{c_{i, \ell}}$  is a backward weak KAM, then  $U_{c_{i, r}}$  will be a forward weak KAM, and vice versa.

When a minimal measure is supported on periodic orbit, of which the lift divides  $\mathbb{R}^2$  into two half plane. The globally elementary weak KAM solution admits a decomposition of periodic solution and an affine function when it is restricted on half plain (see Proposition A.6). If the flat  $\mathbb{F}_{E, i}$  is non-trivial (not a point), these two affine functions are different.

**9.3. Extension of weak KAM solutions using the normal hyperbolicity.** If such modulus continuity is established on certain normally hyperbolic manifold, we can extend it to the stable and unstable fibers.

**Theorem 9.3.** *Let  $\mathbb{T}^k \times \mathbb{R}^k \subset \mathbb{T}^n \times \mathbb{R}^n$  be a NHIM for the Hamiltonian flow and let  $u_\sigma^\pm$  be elementary weak KAMs defined on  $\mathbb{T}^n$ . If  $u_\sigma^\pm|_{\mathbb{T}^k}$  is  $2\kappa$ -Hölder continuous in  $\sigma$ , then  $u_\sigma^\pm$  is  $\kappa$ -Hölder continuous in the parameter  $\sigma$ .*

*Proof.* By definition, certain constant  $C > 0$  exists such that  $|u_\sigma^- - u_{\sigma'}^-| \leq C|\sigma - \sigma'|^{2\kappa}$ . Given two parameters  $\sigma, \sigma'$ , we claim that in each  $|\sigma - \sigma'|^\kappa$ -ball in  $\mathbb{T}^k$  contains a point  $x$  where  $|\partial u_\sigma^-(x) - \partial u_{\sigma'}^-(x)| \leq 4C|\sigma - \sigma'|^\kappa$ . Since Lipschitz function is differentiable almost everywhere, if  $|\partial u_\sigma^- - \partial u_{\sigma'}^-| > 4C|\sigma - \sigma'|^\kappa$  holds for almost every point in the  $|\sigma - \sigma'|^\kappa$ -ball centered at  $x$ , there exists a point  $x'$  in this ball and a path  $\zeta$  connecting these two points such that

$$|(u_\sigma^-(x') - u_{\sigma'}^-(x')) - (u_\sigma^-(x) - u_{\sigma'}^-(x))| = \left| \int_\zeta \langle \partial u_\sigma^- - \partial u_{\sigma'}^-, d\Gamma \rangle \right| > 2C|\sigma - \sigma'|^{2\kappa},$$

which contradicts the  $2\kappa$ -Hölder continuity.

As  $\mathbb{T}^k \times \mathbb{R}^k$  is a normally hyperbolic invariant manifold,  $z = (\partial u_\sigma^-(x), x)$  is a point in this NHIM if  $x$  lies in this  $k$ -torus where  $u_\sigma^-$  is differentiable. This point has its own stable and unstable fiber, denoted by  $\Gamma_z^\pm$ . Because the differential of backward weak KAM determines backward  $c(\sigma)$ -minimal orbit, almost every point  $x'$  in a neighborhood of  $\mathbb{T}^k$  uniquely determines a point  $(x', \partial u_\sigma^-(x'))$  which lies on certain unstable

fiber  $\Gamma_z^-$  with  $z = (\partial u_\sigma^-(x), x)$  and we have

$$u_\sigma^-(x') - u_\sigma^-(x) = \int \langle \partial u_\sigma^-, d\pi \Gamma_z^- \rangle,$$

where  $\pi \Gamma_z^-$  denotes the projection of  $\Gamma_z^-$  down to the configuration space. Since the unstable fiber  $\Gamma_z^-$  is  $C^{r-1}$  continuous in its base point, in  $|\sigma - \sigma'|^\kappa$ -neighborhood of  $x'$  which lies in a neighborhood of  $\mathbb{T}^k$  there exists a point  $x''$  such that both  $u_\sigma^-$  and  $u_{\sigma'}^-$  are differentiable at  $x''$ , the point  $(x'', \partial u_\sigma^-(x''))$  belongs to a unstable fiber  $\Gamma_z^-$  which is  $|\sigma - \sigma'|^\kappa$ -close to another unstable fiber  $\Gamma_{z'}^-$  to which  $(x'', \partial u_{\sigma'}^-(x''))$  belongs. Therefore we obtain

$$\begin{aligned} |u_{\sigma'}^-(x') - u_\sigma^-(x')| &\leq |u_{\sigma'}^-(x'') - u_\sigma^-(x'')| + C_1 |\sigma - \sigma'|^\kappa \\ &\leq |u_{\sigma'}^-(\pi z') - u_\sigma^-(\pi z)| + C_1 |\sigma - \sigma'|^\kappa \\ &\quad + \left| \int \langle \partial u_{\sigma'}^-, d\pi \Gamma_{z'}^- \rangle - \langle \partial u_\sigma^-, d\pi \Gamma_z^- \rangle \right| \\ &\leq C_2 |\sigma - \sigma'|^\kappa, \end{aligned}$$

where the first inequality is due to the Lipschitz property of weak KAM, the third one is obtained from the fact that two fibers keep  $|\sigma - \sigma'|^\kappa$ -close to each other.  $\square$

#### APPENDIX A. A BRIEF INTRODUCTION TO MATHER THEORY AND WEAK KAM

We use the variational method to prove the main result, which is based on Mather theory. In this appendix, we give a brief introduction to the Mather theory and weak KAM theory.

**A.1. Minimizing measure and  $\alpha, \beta$  function.** The theory is established for Tonelli Lagrangian.

**Definition A.1.** Let  $M$  be a closed manifold. A  $C^2$ -function  $L: TM \times \mathbb{T} \rightarrow \mathbb{R}$  is called Tonelli Lagrangian if it satisfies the following conditions:

- (1) *Positive definiteness.* For each  $(x, t) \in M \times \mathbb{T}$ , the Lagrangian function is strictly convex in velocity: the Hessian  $\partial_{\dot{x}\dot{x}} L$  is positive definite.
- (2) *Super-linear growth.* We assume that  $L$  has fiber-wise superlinear growth: for each  $(x, t) \in M \times \mathbb{T}$ , we have  $L/\|\dot{x}\| \rightarrow \infty$  as  $\|\dot{x}\| \rightarrow \infty$ .
- (3) *Completeness.* All solutions of the Lagrangian equations are well defined for the whole  $t \in \mathbb{R}$ .

For autonomous systems, the completeness is automatically satisfied, as each orbit entirely stays in certain compact energy level set.

Given a closed 1-form  $\langle \eta_c(x), dx \rangle$  with first cohomology class  $[\langle \eta_c(x), dx \rangle] = c$ , we introduce a Lagrange multiplier  $\eta_c = \langle \eta_c(x), \dot{x} \rangle$ . Without danger of confusion, we call it closed 1-form also.

For each  $C^1$  curve  $\gamma: \mathbb{R} \rightarrow M$  with period  $k$ , there is unique probability measure  $\mu_\gamma$  on  $TM \times \mathbb{T}$  so that the following holds

$$\int_{TM \times \mathbb{T}} f d\mu_\gamma = \frac{1}{k} \int_0^k f(d\gamma(s), s) ds$$

for each  $f \in C^0(TM \times \mathbb{T}, \mathbb{R})$ , where we use the notation  $d\gamma = (\gamma, \dot{\gamma})$ . Let

$$\mathfrak{H}^* = \{\mu_\gamma \mid \gamma \in C^1(\mathbb{R}, M) \text{ is periodic of } k\}.$$

The set  $\mathfrak{H}$  of holonomic probability measures is the closure of  $\mathfrak{H}^*$  in the vector space of continuous linear functionals. One can see that  $\mathfrak{H}$  is convex.

For each  $\nu \in \mathfrak{H}$  the action  $A_c(\nu)$  is defined as follows

$$A_c(\nu) = \int (L - \eta_c) d\nu.$$

It is proved in [M91, Me] that for each co-homology class  $c$  there exists at least one invariant probability measure  $\mu_c$  minimizing the action over  $\mathfrak{H}$

$$A_c(\mu_c) = \inf_{\nu \in \mathfrak{H}} \int (L - \eta_c) d\nu,$$

called  $c$ -minimal measure. Let  $\mathfrak{H}_c \subset \mathfrak{H}$  be the set of  $c$ -minimal measures, the Mather set  $\tilde{\mathcal{M}}(c)$  is defined as

$$\tilde{\mathcal{M}}(c) = \bigcup_{\mu_c \in \mathfrak{H}_c} \text{supp} \mu_c.$$

The  $\alpha$ -function is defined as  $\alpha(c) = -A_c(\mu_c) : H^1(M, \mathbb{R}) \rightarrow \mathbb{R}$ , it is convex, finite everywhere with super-linear growth. Its Legendre transformation  $\beta : H_1(M, \mathbb{R}) \rightarrow \mathbb{R}$  is called  $\beta$ -function

$$\beta(\omega) = \max_c (\langle \omega, c \rangle - \alpha(c)).$$

It is also convex, finite everywhere with super-linear growth (see [M91]).

Note that  $\int \lambda d\mu_\gamma = 0$  holds for each exact 1-form  $\lambda$  and each  $\mu_\gamma \in \mathfrak{H}^*$ . Therefore, for each measure  $\mu \in \mathfrak{H}$  one can define its rotation vector  $\omega(\mu) \in H_1(M, \mathbb{R})$  such that

$$\langle [\lambda], \omega(\mu) \rangle = \int \lambda d\mu,$$

holds for every closed 1-form  $\lambda$  on  $M$ . By the following relation

$$c \in \mathcal{L}_\beta(\rho) \iff \alpha(c) + \beta(\rho) = \langle c, \rho \rangle.$$

one obtains the Fenchel-Legendre transformation  $\mathcal{L}_\beta : H_1(M, \mathbb{R}) \rightarrow H^1(M, \mathbb{R})$ .

**A.2. (Semi)-static curves, the Aubry set and the Mañé set.** The concept of semi-static curves is introduced by Mather and Mañé (cf. [M93, Me]). A curve  $\gamma : \mathbb{R} \rightarrow M$  is called  $c$ -semi-static if in time-1-periodic case we have

$$[A_c(\gamma)|_{[t, t']}] = F_c((\gamma(t), t), (\gamma(t'), t'))$$

where

$$\begin{aligned} [A_c(\gamma)|_{[t, t']}] &= \int_t^{t'} \left( L(d\gamma(t), t) - \eta_c(d\gamma(t)) \right) dt + \alpha(c)(t' - t), \\ F_c((x, t), (x', t')) &= \inf_{\substack{\tau=t \bmod 1 \\ \tau'=t' \bmod 1}} h_c((x, \tau), (x', \tau')), \end{aligned}$$

in which

$$h_c((x, \tau), (x', \tau')) = \inf_{\substack{\xi \in C^1 \\ \xi(\tau)=x \\ \xi(\tau')=x'}} [A_c(\xi)|_{[\tau, \tau']}].$$

In autonomous case, the period is considered as any positive number. Consequently, the notation of semi-static curve in this case is somehow simpler

$$[A_c(\gamma)|_{(t, t')}] = F_c(\gamma(t), \gamma(t')),$$

where

$$F_c(x, x') = \inf_{\tau > 0} h_c((x, 0), (x', \tau)).$$

**Convention:** Let  $I \subseteq \mathbb{R}$  be an interval (either bounded or unbounded). A continuous map  $\gamma: I \rightarrow M$  is called curve. If it is differentiable, the map  $d\gamma = (\gamma, \dot{\gamma}): I \rightarrow TM$  is called orbit. When the implication is clear without danger of confusion, we use the same symbol to denote the graph,  $\gamma := \cup_{t \in I} (\gamma(t), t)$  is called curve and  $d\gamma := \cup_{t \in I} (\gamma(t), \dot{\gamma}(t), t)$  is called orbit. In autonomous system, the terminology also applies to the image:  $\gamma := \cup_{t \in I} \gamma(t)$  is called curve and  $d\gamma := \cup_{t \in I} (\gamma(t), \dot{\gamma}(t))$  is called orbit.

A semi-static curve  $\gamma \in C^1(\mathbb{R}, M)$  is called  $c$ -static if, in addition, the relation

$$[A_c(\gamma)|_{(t, t')}] = -F_c((\gamma(t'), \tau'), (\gamma(t), \tau))$$

holds in time-1-periodic case and

$$[A_c(\gamma)|_{(t, t')}] = -F_c(\gamma(t'), \gamma(t))$$

holds in autonomous case. An orbit  $X(t) = (d\gamma(t), t \bmod 2\pi)$  is called  $c$ -static (semi-static) if  $\gamma$  is  $c$ -static (semi-static). We call the Mañé set  $\tilde{\mathcal{N}}(c)$  the union of  $c$ -semi-static orbits

$$\tilde{\mathcal{N}}(c) = \bigcup \{d\gamma : \gamma \text{ is } c\text{-semi static}\}$$

and call the Aubry set  $\tilde{\mathcal{A}}(c)$  the union of  $c$ -static orbits

$$\tilde{\mathcal{A}}(c) = \bigcup \{d\gamma : \gamma \text{ is } c\text{-static}\}.$$

We use  $\mathcal{M}(c)$ ,  $\mathcal{A}(c)$  and  $\mathcal{N}(c)$  to denote the standard projection of  $\tilde{\mathcal{M}}(c)$ ,  $\tilde{\mathcal{A}}(c)$  and  $\tilde{\mathcal{N}}(c)$  from  $TM \times \mathbb{T}$  to  $M \times \mathbb{T}$  respectively. They satisfy the inclusion relation

$$\tilde{\mathcal{M}}(c) \subseteq \tilde{\mathcal{A}}(c) \subseteq \tilde{\mathcal{N}}(c).$$

It is showed in [M91, M93] that the inverse of the projection is Lipschitz when it is restricted to  $\mathcal{A}(c)$  as well as to  $\mathcal{M}(c)$ . By adding subscript  $s$  to  $\mathcal{N}$ , i.e.  $\mathcal{N}_s$  we denote its time- $s$ -section. This principle also applies to  $\tilde{\mathcal{N}}(c)$ ,  $\tilde{\mathcal{A}}(c)$ ,  $\tilde{\mathcal{M}}(c)$ ,  $\mathcal{A}(c)$  and  $\mathcal{M}(c)$  to denote their time- $s$ -section respectively. For autonomous systems, these sets are defined without the time component.

On the time-1-section of Aubry set a pseudo-metric  $d_c$  is introduced by Mather in [M93], its definition relies on the quantity  $h_c^\infty$ . Let

$$h_c^\infty((x, s), (x', s')) = \liminf_{\substack{s=t \bmod 1 \\ t'=s' \bmod 1 \\ t'-t \rightarrow \infty}} h_c((x, t), (x', t')),$$

For autonomous system

$$h_c^\infty(x, x') = \liminf_{\tau \rightarrow \infty} h_c((x, 0), (x', \tau)).$$

The pseudo-metric  $d_c$  on Aubry set is defined as

$$d_c((x, t), (x', t')) = h_c^\infty((x, t), (x', t')) + h_c^\infty((x', t'), (x, t)).$$

With the pseudo-metric  $d_c$  one defines equivalence class in Aubry set. The equivalence  $(x, t) \sim (x', t')$  implies  $d_c((x, t), (x', t')) = 0$ , with which one can define quotient Aubry set  $\mathcal{A}(c)/\sim$ . Its element is called Aubry class, denoted by  $\mathcal{A}_i(c)$  or  $\mathcal{A}_{c,i}$ , its lift to  $TM \times \mathbb{T}$  is denoted by  $\tilde{\mathcal{A}}_i(c)$ . Thus,

$$\mathcal{A}(c) = \bigcup_{i \in \Lambda} \mathcal{A}_i(c), \quad \tilde{\mathcal{A}}(c) = \bigcup_{i \in \Lambda} \tilde{\mathcal{A}}_i(c).$$

Although Mather constructed an example with a quotient Aubry set homeomorphic to an interval, it is generic that each  $c$ -minimal measure contains not more than  $n+1$

ergodic components if the system has  $n$  degrees of freedom [BC]. In this case, each Aubry set contains at most  $n + 1$  classes.

The definition of semi-static curve as well as of Mañé set depends on which configuration manifold under our consideration. Let  $\pi : \bar{M} \rightarrow M$  be a finite covering, a curve  $\gamma : \mathbb{R} \rightarrow M$  is said semi-static in  $\bar{M}$  if each curve in its lift  $\bar{\gamma}$  is semi-static in  $\bar{M}$ . Accordingly, we define  $\tilde{\mathcal{N}}(c, \bar{M})$  ( $\mathcal{N}(c, \bar{M})$ ) as the set containing all  $c$ -semi-static orbits (curves) in  $\bar{M}$ . We use the symbol  $\mathcal{N}(c)$  when  $M$  is defaulted as the configuration manifold.

It is possible that  $\pi\mathcal{N}(c, \bar{M}) \supsetneq \mathcal{N}(c, M)$ . For instance, if  $N \subset M$  is a open region such that  $H_1(M, N, \mathbb{Z}) \neq H_1(M, \mathbb{Z})$ ,  $\mathcal{N} \subset N$  and the lift of  $N$  in  $\bar{M}$  has more than one connected component, then this phenomenon takes place. But we have

**Proposition A.1** (Proposition 1.1 of [C12]). *Let  $\pi : \bar{M} \rightarrow M$  be a finite covering space, then*

$$\pi\mathcal{A}(c, \bar{M}) = \mathcal{A}(c, M).$$

**A.3. Elementary Weak KAM.** The concept of  $c$ -semi-static curves can be extended to the curves only defined on  $\mathbb{R}^\pm$ , which are called forward (backward)  $c$ -semi-static curves, denoted by  $\gamma_c^\pm$  respectively. A curve  $\gamma_c^-$  ( $\gamma_c^+$ ) produces a backward (forward) semi-static orbit orbit  $(\gamma_c^-, \dot{\gamma}_c^-)$  ( $(\gamma_c^+, \dot{\gamma}_c^+)$ ).

**Proposition A.2.** *If the Lagrangian  $L$  is Tonelli type, for each point  $(x, \tau) \in M \times \mathbb{T}$ , there is at least one  $\gamma_c^\pm(t, x, \tau)$  which is forward (backward) semi-static curve.*

Since both the  $\omega$ -limit set of  $d\gamma_c^+$  and the  $\alpha$ -limit set of  $d\gamma_c^-$  are in the Aubry set one define

$$W_c^\pm = \bigcup_{(x, \tau) \in M \times \mathbb{T}} \left\{ x, \tau, \frac{d\gamma_c^\pm(\tau, x, \tau)}{dt} \right\},$$

and call  $W_c^+$  the stable set,  $W_c^-$  the unstable set of the  $c$ -minimal measure respectively. If  $\dot{\gamma}^-(\tau, x, \tau) = \dot{\gamma}^+(\tau, x, \tau)$  holds for some  $(x, \tau) \in M \times \mathbb{T}$ , passing through the point  $(x, \tau, \dot{\gamma}_c^-(\tau, x, \tau))$  the orbit is either in the Aubry set or homoclinic to this Aubry set.

If the Aubry set consists of one class, the stable as well as the unstable set has its own generating function  $u_c^\pm$  such that  $W_c^\pm = \text{Graph}(du_c^\pm)$  holds almost everywhere [Fa1]. These functions are weak KAM solutions. We use  $u_c^\pm$  to denote the weak KAM solution for the Lagrangian  $L - \eta_c$ , where  $\eta_c$  is a closed form with  $[\eta_c] = c$ . These functions are Lipschitz, thus differentiable almost everywhere. At each differentiable point  $(x, \tau)$ ,  $(x, \tau, \partial_x u^-(x, \tau))$  uniquely determines backward  $c$ -semi static curve  $\gamma_x^- : (-\infty, \tau] \rightarrow M$  such that  $\gamma_x^-(\tau) = x$ ,  $\dot{\gamma}_x^-(\tau) = \partial_y H(x, \tau, \partial_x u^-(x, \tau))$ . Similarly,  $(x, \tau, \partial_x u^+(x, \tau))$  uniquely determines forward  $c$ -semi static curve  $\gamma_x^+ : [\tau, \infty) \rightarrow M$  such that  $\gamma_x^+(\tau) = x$ ,  $\dot{\gamma}_x^+(\tau) = \partial_y H(x, \tau, \partial_x u^+(x, \tau))$ .

Recall the quantity  $h_c^\infty(x, x')$  for autonomous system. It is a backward weak KAM if we think it as the function of  $x'$ , a forward weak KAM for the variable of  $x$ .

**Elementary weak KAM solution** is introduced when the Aubry set consists of finitely many classes. One Aubry class  $\mathcal{A}_{c,i}$  determines a pair of elementary weak KAM  $u_{c,i}^\pm$ . Let  $x \in \mathcal{A}_{c,i}$ , one has a decomposition

$$h_c^\infty(x, x') = u_{c,i}^-(x') - u_{c,i}^+(x),$$

where  $u_{c,i}^-$  is a backward weak KAM, we call it elementary for  $\mathcal{A}_{c,i}$ . Similarly, one can also define  $u_{c,i}^+$ . We introduce a barrier function for Aubry classes  $\mathcal{A}_{c,i}$  and  $\mathcal{A}_{c,j}$

$$B_{c,i,j} = u_{c,i}^- - u_{c,j}^+.$$

Elementary weak KAM solutions generate all weak KAM solutions in the sense as follows.

**Proposition A.3.** *Assume the minimal measure consists of  $m$  ergodic components. For each weak KAM solution  $u^\pm$ , there exist  $m'$  ( $m' \leq m$ ) constants  $d_1^\pm, \dots, d_{m'}^\pm$  and  $m'$  open domains  $D_1^\pm, \dots, D_{m'}^\pm$  such that they do not overlap each other,  $M = \cup_{1 \leq i \leq m'} \bar{D}_i^\pm$  and*

$$(A.1) \quad u^\pm|_{D_i^\pm} = u_i^\pm + d_i^\pm, \quad \forall 1 \leq i \leq m'.$$

*Proof.* It is deduced from the Lipschitz property of  $u^-$  that it is differentiable almost every where. Let  $x$  be a point where  $u^-$  is differentiable,  $du^-(x)$  determines a unique backward semi static orbit  $d\gamma_c^i: (-\infty, 0] \rightarrow M$  whose  $\alpha$ -limit set is in certain Aubry class  $\tilde{\mathcal{A}}_c^i$ . By definition we have

$$u^-(x) - u^-(\gamma_c^i(-t)) = \int_{-t}^0 L_c(d\gamma_c^i(s), s) ds + \alpha(c)t.$$

Let  $t_k \rightarrow \infty$  such that  $\gamma_c^i(-t_k) \rightarrow x' \in \mathcal{A}_c^i$ , one has

$$(A.2) \quad u^-(x) = h_c^\infty(x', x) + u^+(x') = u_{c,i}^-(x) + d_i,$$

since the difference of any two weak KAMs is constant when they are restricted on an Aubry class.

If  $x^* \in M$  is another point where  $du^-(x^*)$  determines a backward semi-static orbits whose  $\alpha$ -limit set is also contained in  $\tilde{\mathcal{A}}_c^i$ , we then obtain (A.2) for  $u^-(x^*)$  with the same  $d_i$ . All these points constitute a set connected with  $\mathcal{A}_c^i$ . There are not more than  $m$  connected sets such that (A.1) holds.  $\square$

**Globally elementary weak KAM solution** is defined for the universal covering space of the configuration manifold. It is achieved by considering the elementary weak KAM solutions for arbitrary finite covering space of  $\mathbb{T}^n$ .

Let  $\mathbf{k} = (k_1, \dots, k_n) \in \mathbb{Z}^n$  with  $k_i \geq 1$ ,  $\mathbf{k}\mathbb{T}^n = \{x \in \mathbb{R}^n : x_i \bmod k_i\}$ . Given a finite covering  $\pi_{\mathbf{k}}: \mathbf{k}\mathbb{T}^n \rightarrow \mathbb{T}^n$ , a function  $u_c: \mathbb{T}^n \rightarrow \mathbb{R}$  has its natural lift  $\bar{u}_{c,\mathbf{k}}: \mathbf{k}\mathbb{T}^n \rightarrow \mathbb{R}$  in the way that  $\bar{u}_{c,\mathbf{k}}(\bar{x}) = u_c(\pi_{\mathbf{k}}\bar{x})$ . If a finite covering of  $\mathbb{T}^n$  is considered to be configuration space, the weak KAM solution may not be  $2\pi$ -periodic. The lift of elementary weak KAM solution may not be elementary. However, one has

**Proposition A.4.** *If there is only one Aubry class for the finite covering  $\pi_{\mathbf{k}}: \mathbf{k}\mathbb{T}^n \rightarrow \mathbb{T}^n$ , then the globally elementary weak KAM solution, denoted by  $u_{c,\mathbf{k}}^\pm$ , is the lift of the elementary weak KAM solution for  $\mathbb{T}^n$ :  $u_{c,\mathbf{k}}^\pm = \bar{u}_{c,\mathbf{k}}^\pm$ .*

*Proof.* For a finite covering  $\pi_{\mathbf{k}}: \mathbf{k}\mathbb{T}^n \rightarrow \mathbb{T}^n$ , if the Aubry set still consists of only one class, then only one pair of weak KAM solution exists, denoted by  $u_{c,\mathbf{k}}^\pm$ .

Let  $d_{c,\mathbf{k}}, h_{c,\mathbf{k}}^\infty$  be the quantity of  $d_c, h_c^\infty$  defined for the configuration space  $\mathbf{k}\mathbb{T}^n$ , then  $h_{c,\mathbf{k}}^\infty(\bar{x}, \bar{x}') \geq h_c^\infty(\pi_{\mathbf{k}}\bar{x}, \pi_{\mathbf{k}}\bar{x}') \geq 0$  if  $\pi_{\mathbf{k}}\bar{x} = \pi_{\mathbf{k}}\bar{x}'$ . As  $d_c(\bar{x}, \bar{x}') = h_{c,\mathbf{k}}^\infty(\bar{x}, \bar{x}') + h_{c,\mathbf{k}}^\infty(\bar{x}', \bar{x}) = 0$



if  $\pi_k \bar{x} = \pi_k \bar{x}' \in \mathcal{A}_c$ , it follows that  $h_{c,k}^\infty(\bar{x}, \bar{x}') = 0$ . Consequently,  $u_{c,k}^\pm(\bar{x}) = u_{c,k}^\pm(\bar{x}')$  if  $\pi_k \bar{x} = \pi_k \bar{x}' \in \mathcal{A}_c$ .

Let  $D_k$  be a Deck transformation on  $k\mathbb{T}^n$ . For any  $\bar{x}, \bar{x}' \in k\mathbb{T}^n$  with  $\pi_k \bar{x}' \in \mathcal{A}_c$ , one has

$$\begin{aligned} u_{c,k}^-(\bar{x}) - u_{c,k}^+(\bar{x}') &= h_{c,k}^\infty(\bar{x}, \bar{x}') = h_{c,k}^\infty(D_k \bar{x}, D_k \bar{x}') \\ &= u_{c,k}^-(D_k \bar{x}) - u_{c,k}^+(D_k \bar{x}') \\ &= u_{c,k}^-(D_k \bar{x}) - u_{c,k}^+(\bar{x}'). \end{aligned}$$

Therefore, one has  $u_{c,k}^-(\bar{x}) = u_{c,k}^-(D_k \bar{x})$  holds for any Deck transformation on  $k\mathbb{T}^n$ .  $\square$

The lift of an Aubry class  $\mathcal{A}_i$  may contain several connected components  $\{\mathcal{A}_{i,j}\}$ . It is closely related to whether the first relative homology of  $\mathbb{T}^n$  with respect to  $\mathcal{A}_i$  is trivial. Let

$$\mathbb{H}_i = \bigcap_U \{i_{U*} H_1(U, \mathbb{R}) : U \text{ is a neighborhood of } \mathcal{A}_i \text{ in } \mathbb{T}^n\},$$

where  $i_U: U \rightarrow M$  denotes the inclusion map. Small neighborhood  $U$  exists such that  $\mathbb{H}_i = i_{U*} H_1(U, \mathbb{R})$ . If  $H_1(\mathbb{T}^n, \mathbb{Z}) \supsetneq \mathbb{H}_i$ , suitable finite covering spaces exist such that the lift of an Aubry class contains more than one connected component.

For  $k\mathbb{T}^n$  with large  $\min k_j$ , we fix one Aubry class, denoted by  $\mathcal{A}_i^k$ . It is distinguished by fixing a point  $x \in \{x \in k\mathbb{T}^n : 0 \leq x_i < 1\} \cap \mathcal{A}_i$  such that  $x \in \mathcal{A}_i^k$ , which determines a pair of elementary weak KAM  $u_{i,k}^\pm$ . The elementary weak KAM solution of other Aubry class in the lift is just a shift of this function  $u_{i,k}^\pm(x) \rightarrow u_{i,k}^\pm(x + x_0)$ .

**Proposition A.5** (Lemma 2.1 of [C12]). *Given an Aubry class  $\mathcal{A}_i$  and a bounded domain  $\Omega \subset \mathbb{R}^n$ , there exists  $R_\Omega > 0$  such that for any  $k, k'$  with  $\min\{k_i, k'_i\} \geq R_\Omega$  such that*

$$u_{i,k}^\pm|_\Omega = u_{i,k'}^\pm|_\Omega + \text{constant}$$

where  $u_{i,k}^\pm, u_{i,k'}^\pm$  are treated as  $2k\pi$ -periodic and  $2k'\pi$ -periodic function respectively.

Therefore, some function  $u_{i,\infty}^\pm$  is well-defined (called globally elementary weak KAM solution) satisfying the condition: given any bounded domain  $\Omega$ , there exists  $R_\Omega > 0$  such that

$$u_{i,\infty}^\pm|_\Omega = u_{i,k}^\pm|_\Omega + \text{constant}, \quad \text{provided } \min\{k_i\} \geq R_\Omega.$$

The globally elementary weak KAM solution has some special property if the Aubry set contains a connected component homeomorphic to  $\mathbb{T}^{n-1}$ . In this case, the lift of  $\mathcal{A}_i$  divides  $\mathbb{R}^n$  into two parts, denoted by  $R^-$  and  $R^+$ .

**Proposition A.6** (Theorem 2.5 of [C12]). *If the Aubry set contains a connected component contains an Aubry class  $\mathcal{A}_i$  homeomorphic to  $\mathbb{T}^{n-1}$ , then the globally elementary weak KAM solution  $u_{i,\infty}^\pm: \mathbb{R}^n \rightarrow \mathbb{R}$  has a decomposition*

$$u_{i,\infty}^\pm = v_{i,\infty}^\pm + w_{i,\infty}^\pm,$$

where  $v_{i,\infty}^\pm$  is periodic and  $w_{i,\infty}^\pm$  is affine when they are restricted in the half space  $R^+$  as well as in another half space  $R^-$ .

For a convex function one can define its sub-derivative. The set of sub-derivative of a convex function  $\psi$  at  $x$

$$D^-\psi(x) = \{y \in \mathbb{R}^n : \psi(x') - \psi(x) \geq \langle y, x' - x \rangle \quad \forall x' \in \mathbb{R}^n\}$$

is convex. Since each backward weak KAM solution  $u^-$  has a decomposition  $u^- = \phi - \psi$  where  $\phi$  is smooth and  $\psi$  is convex, we define

$$(A.3) \quad D^+u^-(x) = \{\phi'(x) - y : y \text{ is a sub-derivative of } \psi \text{ at } x\}.$$

**Theorem A.1.** *Let  $u_c^-$  be a backward weak KAM,  $\gamma_c : (-\infty, 0] \rightarrow M$  is  $(u_c^-, L_c)$ -calibrated curve. Then,  $\frac{\partial L_c}{\partial \dot{x}}(\gamma_c(0), \dot{\gamma}_c(0)) + c$  holds on the boundary of  $D^+u_c^-(x)$ . In particular,  $y$  is a vertex of  $D^+u_c^-(x)$  only when there exists a calibrated curve  $\gamma$  such that  $y + c = \frac{\partial L_c}{\partial \dot{x}}(\gamma_c(0), \dot{\gamma}_c(0))$ .*

*Proof.* A curve  $\gamma : (-\infty, 0] \rightarrow M$  is called  $(u, L_c)$ -calibrated if

$$u(\gamma(t')) - u(\gamma(t)) = \int_t^{t'} L_c(\gamma(s), \dot{\gamma}(s)) ds + (t' - t)\alpha(c)$$

holds for each  $-\infty < t \leq t' \leq 0$ . We say that  $(x, v)$  determines a  $(u_c^-, L_c)$ -calibrated curve  $\gamma : (-\infty, 0] \rightarrow M$  if  $\gamma(0) = x$  and  $\dot{\gamma}(0) = v$ . We say that  $v \in Vu_c^-(x) \subset \mathbb{R}^n$  if  $(x, v)$  determines certain  $(u_c^-, L_c)$ -calibrated curve  $\gamma : (-\infty, 0] \rightarrow M$ . By Proposition 4.5.1 in [Fa2], one has

$$(A.4) \quad u_c^-(x + \delta h) - u_c^-(x) \leq \delta \left\langle \frac{\partial L_c}{\partial \dot{x}}(x, v), h \right\rangle + O(\delta^2)$$

for each  $(v, h) \in Vu_c^-(x) \times \mathbb{R}^n$ . It implies that

$$\text{co}\{\partial_{\dot{x}}L_c(x, v) : v \in Vu_c^-(x)\} \subseteq D^+u_c^-(x).$$

We use  $\text{co}S$  denote the convex hull of the set  $S$ .

Adding a constant to the Lagrangian  $L_c \rightarrow L_c + \alpha(c)$ , we assume  $\alpha(c) = 0$ . Also, we only need to consider the case  $c = 0$ . Let  $\gamma : (-\infty, 0] \rightarrow M$  be a backward  $c$ -semi-static curve, calibrated for  $u_c^-$ , with  $\gamma(0) = x$  and  $\dot{\gamma}(0) = v$ . Let  $x_1 = \gamma(-\delta)$  with  $\delta > 0$ , then  $x - x_1 = \delta v + O(\delta^2)$  if  $\delta$  is sufficiently small. Let  $\mu = \delta + \nu$  with  $\nu > 0$ ,  $\xi : [-\mu, -\delta] \rightarrow M$  be a curve connecting  $\gamma(-\mu)$  to  $\gamma(0)$  such that  $\xi(s) = \gamma(\frac{\nu+\delta}{\nu}(s + \delta))$

$$(A.5) \quad \begin{aligned} u^-(x) - u^-(\gamma(-\mu)) &= \int_{-\mu}^0 L(\gamma(s), \dot{\gamma}(s)) ds \\ &\leq \int_{-\mu}^{-\delta} L(\xi(s), \dot{\xi}(s)) ds. \end{aligned}$$

$$(A.6) \quad u^-(x_1) - u^-(\gamma(-\mu)) = \int_{-\mu}^{-\delta} L(\gamma(s), \dot{\gamma}(s)) ds,$$

As  $\delta > 0$  is assumed small, by definition we obtain

$$\begin{aligned} \|\gamma(s) - \xi(s)\| &= O(|\delta|), \\ \|\dot{\gamma}(s) - \dot{\xi}(s)\| &= \frac{\delta + \nu}{\nu} O(|\delta|), \quad \forall s \in [-\mu, -\delta]. \end{aligned}$$

Since  $\gamma(s)$  is a solution of the Lagrange equation,  $\partial_x L(\gamma(s), \dot{\gamma}(s)) = \frac{d}{ds} \partial_{\dot{x}} L(\gamma(s), \dot{\gamma}(s))$ . Consequently, one has

$$\begin{aligned} L(\gamma(s), \dot{\gamma}(s)) - L(\xi(s), \dot{\xi}(s)) &= \frac{d}{ds} \left\langle \frac{\partial L}{\partial \dot{x}}(\gamma(s), \dot{\gamma}(s)), \gamma(s) - \xi(s) \right\rangle \\ &\quad + \left( \frac{\delta + \nu}{\nu} \right)^2 O(\delta^2). \end{aligned}$$

Since  $\gamma(-\delta) = \gamma(0) - \delta v + O(|\delta|^2)$ , we find  $|u^-(x_1) - u^-(x - \delta v)| = O(\delta^2)$  as weak KAM solution is Lipschitz function. Subtracting (A.6) from (A.5), let  $\nu = O(1)$ , note  $\gamma(-\delta) = x_1$ ,  $\xi(-\delta) = x$  and  $\gamma(-\mu) = \xi(-\mu)$ , one deduces

$$\begin{aligned} u^-(x - \delta v) - u^-(x) &\geq \int_{-\mu}^{-\delta} L(\gamma(s), \dot{\gamma}(s)) - L(\xi(s), \dot{\xi}(s)) ds \\ &\geq \left\langle \frac{\partial L}{\partial \dot{x}}(x, v), -\delta v \right\rangle - O(\delta^2). \end{aligned}$$

By using (A.4) one then deduces that

$$u^-(x) - u^-(x - \delta v) = \delta \left\langle \frac{\partial L}{\partial \dot{x}}(x, v), v \right\rangle + O(\delta^2).$$

That is,  $\partial_{\dot{x}} L(x, v)$  is on the boundary of  $D^+ u^-(x)$  provided  $v \in Vu^-(x)$ .

To complete the proof, let us recall the concept of reachable gradient (cf. [CS]):

**Definition A.2.** Let  $u: A \rightarrow \mathbb{R}$  be locally Lipschitz. A vector  $p$  is called a reachable gradient of  $u$  at  $x \in A$  if a sequence  $\{x_k\} \subset A \setminus \{x\}$  exists such that  $u$  is differentiable at  $x_k$  for each  $k \in \mathbb{N}$  and

$$\lim_{k \rightarrow \infty} x_k = x, \quad \lim_{k \in \infty} Du(x_k) = p.$$

The set of all reachable gradient of  $u$  at  $x$  is denoted by  $D^* u(x)$ .

If  $u$  is a semi-concave function then it is proved in [CS] that

$$D^+ u(x) = \text{co} D^* u(x).$$

Therefore, there exists a sequence  $x_k \rightarrow x$  such that  $u^-(x_k)$  is differentiable and  $\partial u^-(x_k) \rightarrow y$  provided  $y$  is a vertex of  $D^+ u^-(x)$ . It is well-known that  $u^-$  is differentiable at  $x$  if and only there is only one  $v \in Vu^-(x)$  such that  $(x, v)$  determines a unique  $(u^-, L)$ -calibrated curve  $\gamma: (-\infty, 0] \rightarrow M$  with  $\gamma(0) = x$  and  $\dot{\gamma}(0) = v$ . The derivative of  $u^-$  at this point is a singleton

$$Du^-(x) = \frac{\partial L}{\partial \dot{x}}(x, v).$$

Let  $x_k \rightarrow x$  be a sequence such that  $u^-$  is differentiable at each  $x_k$ , corresponding speed is denoted by  $v_k$ . If  $v$  is the accumulation point of  $\{v_k\}$ , by the upper semi-continuity of the set of backward (forward) semi-static curves in cohomology class,  $(x, v)$  determines a  $(u^-, L)$ -calibrated curve. Clearly, one has  $y = \frac{\partial L}{\partial \dot{x}}(x, v)$ , which completes the proof.  $\square$

**Acknowledgment.** The first author is supported by some grants from China, such as National Basic Research Program (2013CB834100), NNSF (Grant 10531050) and a program PAPD of Jiangsu Province. The second author is supported by NSF grant DMS-1500897. He would like to thank the first author, Mather and Kaloshin from

whom he learned Arnold diffusion, his mentor Wilkinson for her constant encouragement and support and thank Polterovich for his generous comments. We wish to thank Kaloshin for pointing out an error in the previous version of the manuscript.

## REFERENCES

- [A63] Arnol'd V.I., *Small denominators and problems of stability of motion in classical and celestial mechanics*, Russ. Math. Survey **18** (1963) 85-192.
- [A64] Arnol'd, V.I. *Instability of dynamical systems with several degrees of freedom (Instability of motions of dynamic system with five-dimensional phase space)*. Soviet Mathematics 5 (1964): 581-585.
- [A66] Arnol'd, V.I. *The stability problem and ergodic properties for classical dynamical systems*. Proceedings of International Congress of Mathematicians, Moscow (1966) 5-11, in V. I. Arnold- Collected Works. Springer Berlin Heidelberg (2014) 107-113.
- [A89] Arnol'd, V.I. *Mathematical methods of classical mechanics*. Springer, 1989.
- [A94] Arnol'd, V.I. *Mathematical problems in classical physics*. Trends and perspectives in applied mathematics, 1-20, Appl. Math. Sci., 100, Springer, New York, 1994.
- [ABC] Arnaud, Marie-Claude, Christian Bonatti, and Sylvain Crovisier. *Dynamiques symplectiques g n riques*. Ergodic Theory and Dynamical Systems 25.05 (2005): 1401-1436.
- [ACW] Avila, Artur, Sylvain Crovisier, and Amie Wilkinson. *Diffeomorphisms with positive metric entropy*. arXiv preprint arXiv:1408.4252 (2014).
- [Be1] Bernard P., *Symplectic aspects of Mather theory*, Duke Math. J. **136** (2007) 401-420.
- [Be2] Bernard P., *The dynamics of pseudographs in convex Hamiltonian systems*, Journal Amer. Math. Soc. **21** (2008) 615-669.
- [Be3] Bernard P., *Large normally hyperbolic cylinders in a priori stable Hamiltonian systems*. Annales Henri Poincar  Vol. **11**. No. 5. SP Birkhauser Verlag Basel, 2010.
- [Be4] Bernard, Patrick. *Existence of  $C^{1,1}$  critical sub-solutions of the Hamilton-Jacobi equation on compact manifolds*. Annales scientifiques de l'Ecole normale sup rieure. Vol. 40. No. 3. 2007.
- [BC] Bernard P. & Contreras G., *A generic property of families of Lagrangian systems*, Annals of Math. **167** (2008) 1099-1108.
- [BoCr] Bonatti, Christian, and Sylvain Crovisier. *R currence et g n ricit *. Inventiones mathematicae 158.1 (2004): 33-104.
- [BFRV] Bessa, Mario, Celia Ferreira, Jorge Rocha, and Paulo Varandas. *Generic Hamiltonian Dynamics*. Journal of Dynamics and Differential Equations (2015): 1-16.
- [BKZ] Bernard P. Kaloshin V. & Zhang K., *Arnold diffusion in arbitrary degrees of freedom and crumpled 3-dimensional normally hyperbolic invariant cylinders*. arXiv:1112.2773 (2011).
- [BLZ] Bates P., Lu. K. and Zeng C., *Persistence of overflowing manifolds for semiflow*, Commun. Pure Appl. Math. **52** (1999) 983-1046.
- [Ca] Carneiro, MJ Dias. *On minimizing measures of the action of autonomous Lagrangians*. Non-linearity 8.6 (1995): 1077.
- [CS] Cannarsa P. & Sinestrari C., *Semiconcave functions, Hamilton-Jacobi equations and Optimal Control*, Progress in Nonlinear Differential Equations and Their Applications **58** (2004) Birkh user.
- [CY1] Cheng C.-Q. & Yan J., *Existence of diffusion orbits in a priori unstable Hamiltonian systems*, J. Differential Geometry, **67** (2004) 457-517.
- [CY2] Cheng C.-Q. & Yan J., *Arnold diffusion in Hamiltonian Systems: a priori Unstable Case*, J. Differential Geometry, **82** (2009) 229-277.
- [C11] Cheng C.-Q. *Non-existence of KAM torus*, Acta Math. Sinica, English Series, **27** (2011) 397-404.
- [C12] Cheng C.-Q. *Arnold diffusion in nearly integrable Hamiltonian systems*. arXiv:1207.4016v2 (2013).
- [C15a] Cheng C.-Q. *Uniform hyperbolicity of invariant cylinder*. arXiv:1509.03160 (2015).
- [C15b] Cheng C.-Q. *A way to cross double resonance*. arXiv:1510.08754 (2015).
- [CZ1] Cheng C.-Q. & Zhou M., *Hyperbolicity of minimal periodic orbits*, arXiv:1302.4889 (2013), to appear in Math. Res. Lett.
- [CZ2] Cheng C.-Q. & Zhou M., *Non-degeneracy of extremal points in multi-dimensional space*, Science China Math. **58**(11) (2015) 2255-2260.
- [CW] Cheng C.-Q. & Wang L., *Destruction of Lagrangian torus for positive definite Hamiltonian systems*, GAFA (Geometric and Functional Analysis) **23** (2013) 848-866.

- [DH] Delshams A. & Hugué G., *Geography of resonances and Arnold diffusion in a priori unstable Hamiltonian systems*. Nonlinearity **22.8** (2009): 1997.
- [DLS00] Delshams A. de la Llave R. & Seara T.M. *A Geometric Approach to the Existence of Orbits with Unbounded Energy in Generic Periodic Perturbations by a Potential of Generic Geodesic Flows of  $\mathbb{T}^2$* . Commun. Math. Phys. **209** (2000): 353-392.
- [DLS06] Delshams A. de la Llave R. & Seara T.M. *A geometric mechanism for diffusion in Hamiltonian systems overcoming the large gap problem: heuristics and rigorous verification on a model*. Mem. Amer. Math. Soc **179** (2006): 844
- [DLS08] Delshams A. de la Llave R. & Seara T.M. *Geometric properties of the scattering map of a normally hyperbolic invariant manifold*. Advances in Mathematics 217.3 (2008): 1096-1153.
- [DLS13] Delshams A. de la Llave R. & Seara T.M. *Instability of high dimensional hamiltonian systems: Multiple resonances do not impede diffusion*. arXiv preprint arXiv:1306.4614 (2013).
- [CEL] Crandall M. Evans L.C. and Loins P.-L., *Some properties of viscosity solutions of Hamilton-Jacobi equations*, Trans. Amer. Math. Soc. **282** (1984), 487-502.
- [Fa1] Fathi A., *Théorème KAM faible et théorie de Mather sur les systèmes*, C. R. Acad. Sci. Paris Sér. I Math. **324** (1997) 1043-1046.
- [Fa2] Fathi A., *Weak KAM Theorem in Lagrangian Dynamics*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, (2009).
- [Fe] Fenichel, Neil. *Persistence and smoothness of invariant manifolds for flows*. Indiana Univ. Math. J 21.193-226 (1971): 1972.
- [FS] Fathi A. and Siconolfi A., *Existence of  $C^1$  critical subsolutions of the Hamilton-Jacobi equation*, Invent. Math. **155** (2004) 363-388.
- [GT1] Gelfreich, Vassili, and Dmitry Turaev. *Unbounded energy growth in Hamiltonian systems with a slowly varying parameter*. Communications in Mathematical Physics 283.3 (2008): 769-794.
- [GT2] Gelfreich, Vassili, and Dmitry Turaev. *Arnold Diffusion in a priori chaotic Hamiltonian systems*. arXiv preprint arXiv:1406.2945 (2014).
- [HPS] Hirsch M. Pugh C. and Shub M., *Invariant Manifolds*, Lecture Notes Math. **583** (1977) Springer-Verlag.
- [Is] Ishii H., *A simple, direct proof of uniqueness for solutions of the Hamilton-Jacobi equations*, Proceedings Amer. Math. Soc., **100** (1987), 247-251.
- [KL] Kaloshin, Vadim, and Mark Levi. *Geometry of Arnold diffusion*. SIAM review 50.4 (2008): 702-720.
- [KZ1] Kaloshin V. & Zhang K., *A strong form of Arnold diffusion for two and a half degrees of freedom*. arXiv:1212.1150v2 (2013).
- [KZ2] Kaloshin V. & Zhang K., *A strong form of Arnold diffusion for three and a half degrees of freedom*, (Announcement of result) Preprint (2014). <http://www2.math.umd.edu/~vkaloshi/>, <http://www.math.utoronto.ca/kzhang/publication.html>.
- [KZ3] Kaloshin, V. & Zhang K., *Dynamics of the dominant Hamiltonian, with applications to Arnold diffusion*. arXiv:1410.1844v2 (2015).
- [KZ4] Kaloshin, V., and K. Zhang. *Arnold diffusion for smooth convex systems of two and a half degrees of freedom*. Nonlinearity 28.8 (2015): 2699.
- [L] de la Llave, R. *Orbits of unbounded energy in perturbations of geodesic flows by periodic potentials. A simple construction*. Preprint (2004).
- [Lx] Li X. *On  $c$ -equivalence*, Science in China, series A, **52** (2009) 2389-2396.
- [LC] Li X. & Cheng C-Q., *Connecting orbits of autonomous Lagrangian systems*, Nonlinearity **23** (2009) 119-141.
- [MaSa] Marco, Jean-Pierre, and Lara Sabbagh. *Examples of nearly integrable systems on  $\mathbb{A}^3$  with asymptotically dense projected orbits*. arXiv preprint arXiv:1401.3593 (2014).
- [Mc] Marco, Jean-Pierre. *Twist maps and Arnold diffusion for diffeomorphisms*. conference paper 2015.
- [M91] Mather J., *Action minimizing invariant measures for positive definite Lagrangian systems*, Math. Z., **207(2)**(1991) 169-207.
- [M93] Mather J., *Variational construction of connecting orbits*, Ann. Inst. Fourier (Grenoble), **43(5)**(1993) 1349-1386.
- [M04] Mather J., *Arnold diffusion, I: Announcement of results*, J. Mathematical Sciences, **124(5)** (2004) 5275-5289. (Russian translation in Sovrem. Mat. Fundam. Napravl, **2** (2003) 116-130).
- [M11] Mather, John N. *Shortest curves associated to a degenerate Jacobi metric on  $\mathbb{T}^2$* . Progress in variational methods (2011): 126-168.

- [M12] Mather, John N. *Arnold diffusion by variational methods*. Essays in Mathematics and its Applications. Springer Berlin Heidelberg, 2012. 271-285.
- [Me] Mañé R., *Generic properties and problems of minimizing measures of Lagrangian systems*, Nonlinearity **9** (1996) 273-310.
- [Ms] Massart D., *On Aubry sets and Mather's action functional*, Israel J. Math. **134** (2003) 157-171.
- [MS] McDuff D. and Salamon D., *Introduction to symplectic topology*. (1995).
- [NP] Meysam Nassiri, Enrique R. Pujals *Robust Transitivity in Hamiltonian Dynamics*, Annales scientifiques de l'ENS 45, fascicule 2 (2012), 191-239
- [Os] Osuna O., *Vertices of Mather's beta function*, Ergod. Theory Dynam. Syst. **25** (2005) 949-955.
- [Ri] Rifford L., *On viscosity solutions of certain Hamilton-Jacobi equations: regularity results and generalized Sard's theorem*, Commun. Partial Diff. Eqns. **33** (2008) 517-559.
- [Tr] Treschev D.V., *Evolution of slow variables in a priori unstable Hamiltonian systems*, Nonlinearity, **17** (2004) 1803-1841.
- [Zhe] Zheng Y. & Cheng C.-Q., *Homoclinic orbits of positive definite Lagrangian systems*, J. Diff. Eqns. **229** (2006) 297-316.
- [Zho1] Zhou M., *Hölder regularity of barrier functions in a priori unstable case*, Math. Res. Lett., **18** (2011) 77-94.
- [Zho2] Zhou M., *Infinity of minimal homoclinic orbits*, Nonlinearity, **24** (2011) 931-939.

DEPARTMENT OF MATHEMATICS, NANJING UNIVERSITY, NANJING 210093, CHINA

*E-mail address:* chengcq@nju.edu.cn.

DEPARTMENT OF MATHEMATICS, THE UNIVERSITY OF CHICAGO, CHICAGO, IL60637, USA

*E-mail address:* jxue@math.uchicago.edu.